Noname manuscript No.
(will be inserted by the editor)

Visual Analysis of Pressure in Football

Gennady Andrienko · Natalia Andrienko · Guido Budziak · Jason Dykes · Georg Fuchs · Tatiana von Landesberger · Hendrik Weber

Received: 29.02.2016 / Accepted: hopefully soon

Abstract Modern movement tracking technologies enable acquisition of high quality data about movements of the players and the ball in the course of a football match. However, there is a big difference between the raw data and the insights into team behaviors that analysts would like to gain. To enable such insights, it is necessary first to establish relationships between the concepts characterizing behaviors and what can be extracted from data. This task is challenging since the concepts are not strictly defined. We propose a computational approach to detecting and quantifying the relationships of pressure emerging during a game. Pressure is exerted by defending players upon the ball and the opponents. Pressing behavior of a team consists of multiple instances of pressure exerted by the team members. The extracted pressure relationships can be analyzed in detailed and summarized forms with the use of static and dynamic visualizations and interactive query tools. To support examination of team tactics in different situations, we have designed and implemented a novel interactive visual tool "time mask". It enables selection of multiple disjoint time intervals in which given conditions are fulfilled. Thus, it is possible to select game situations according to ball possession, ball dis-

G. and N. Andrienko
Fraunhofer Institute IAIS, Germany and City University London, UK
E-mail: gennady/natalia.andrienko@iais.fraunhofer.de

G. Budziak
TU Eindhoven, The Netherlands

J. Dykes
City University London, UK

G. Fuchs
Fraunhofer Institute IAIS, Germany

T. von Landesberger
TU Darmstadt, Germany

H. Weber
DFL Deutsche Fussball Liga GmbH, Germany

tance to the goal, time that has passed since the last ball possession change or remaining time before the next change, density of players' positions, or various other conditions. In response to a query, the analyst receives visual and statistical summaries of the set of selected situations and can thus perform joint analysis of these situations. We give examples of applying the proposed combination of computational, visual, and interactive techniques to real data from games in the German Bundesliga, where the teams actively used pressing in their defense tactics.

Keywords Football data analysis · Movement data analysis · Collective movement patterns · Visual Analytics

1 Introduction

Intensive progress in technologies for position tracking has enabled acquisition of high quality data describing the movements of players in sport games, in particular, football (a.k.a. soccer). For ball games, the positions of the players and the ball are measured and recorded at regular time intervals with high temporal resolution (10-25 Hz). Additionally to position data (trajectories), data sets describing game events are created. These may include goals, shots, substitutions, tackles, offsides, and passes between players. While trajectories are usually acquired automatically, event data are created by manual annotation, sometimes supported by computational techniques that suggest candidate game episodes. Additionally, expert annotations about formations used by teams and roles of individual players can be attached to general data about the game. Several companies (e.g. OPTA[1] and ProZone[2]) are active in this market, serving major football leagues and international competitions. Websites and mobile apps like StatZone[3] by OPTA provide interactive visualizations of game statistics, events, player influence etc. Such visualizations mostly rely on manually annotated events, only partly utilizing the potential of positional data. Moreover, these visualizations focus on activities of individual players but don't show their coordination and cooperation within a team and interaction with the opponent team.

Availability of position and event data for football motivated intensive research on football game analysis. Proceeding from elementary movements of individual players reflected in data, researchers have striven to characterize the strategies of the teams and relate characteristics of the movements to team performance in terms of game events and results. Thus, an emerging body of work has focused on spatial arrangements (formations) of groups of players from the same team and correspondences between formations applied by two teams. Another strand of work analyzes the use of the pitch space by teams and amount of space controlled by each team. Many researchers propose

[1] http://www.optasports.com
[2] http://prozonesports.stats.com
[3] http://www.fourfourtwo.com/statszone

approaches to computational synthesis of high-level features that can be used for characterization and classification of game episodes or team play.

A phenomenon that has not yet received much attention from researchers in football data analysis is *pressing*. This is an important and growing trend in modern football tactics. Pressing may be applied by the defending team against the team that possesses the ball. The goal is to win the ball or at least deprive the opponents of the opportunities to develop an attack. *Counter-pressing* is a term denoting a special case of pressing: attempting to re-gain the ball as soon as it has been lost. Effective pressing has been key to the success of several high-level football clubs, such as FC Barcelona, Borussia Dortmund, and Atletico Madrid. Club and media analysts are very interested in getting insights into the way in which teams organize and conduct pressing activity, and the circumstances in which this tactic seems to be effective.

Historically, it was difficult to analyze pressing because it requires data concerning ball possession and movement, while acquisition of ball-related data of sufficient quality has become possible only recently. Another obstacle to developing approaches to analyzing and understanding pressing behaviors is that the notion of pressing as such is ill-defined. While there are many articles in the media talking about pressing and giving examples of pressing, a clear definition of pressing is still lacking. This makes it difficult to find appropriate numeric measures that could be used for detecting pressing situations from position data and characterizing these situations.

In our work, we propose to approximate the ill-defined notion of pressing using a working concept of "pressure". Our idea is that the complex phenomenon of pressing behavior consists of multiple instances of "pressure" exerted by players of the defending team on the ball and on the players of the team possessing the ball. The level of pressure can be numerically estimated based on the relative position of a defense player with respect to the target of the pressure (i.e., the ball or an opponent player) and the defended goal. To estimate the pressure on an opponent player, the position of the ball is also taken into account. We propose a parametrized formula for numeric expression of the level of pressure. We anticipate that this approach will be helpful to football specialists, who can experiment with the parameters to empirically find suitable settings and, possibly, come to a clearer definition of the concept of pressing as captured by high resolution football data.

Our research team is composed of five computer scientists and two football experts, one of them being a former professional player in the Dutch league and the other a sport scientist working for the German Bundesliga. The research task of detecting and analyzing pressing comes from the football experts, who also provided their judgments of ideas, approaches, and results in the course of the work. The resulting achievements and contributions can be summarized as follows:

- **numeric measures of pressure** on the ball and opponents, which enable detection and analysis of pressing behaviors;

- interactive visual tools for **finding episodes that involve pressing** and for viewing them dynamically;
- interactive query (filtering) tools for **extracting game episodes of interest** based on computed measures, for example, pressure by or against a particular team, pressure in the first two seconds after losing the ball, pressure close to the opponents' goal, etc.;
- methods for **summarization of extracted episodes** and techniques for visual representation of the summaries providing answers to questions: Who? When? and Where?
- coordinate transformations and **relative spaces** [7] that capture the distribution of pressing activities across a team and between teams involved in a match.

The data used in this paper consist of trajectories of the players and the ball. The positions were measured and recorded at the temporal resolution of 25Hz, i.e., the time interval between consecutive position records is 40 millisecond. For each position of the ball, there are two additional binary attributes: ball status (in play or out of play) and ball possession (one of the two teams). For each game, about 3,500,000 positions of players and the ball are recorded. Such high quality data from several domestic games from the German Bundesliga were provided by DFL Deutsche Fußball Liga GmbH[4].

The remainder of the paper is structured as follows. In Section 2, we overview related papers on football analytics and movement analysis in general. Section 3 describes our approach to the quantification of pressure, demonstrates visual displays of game episodes involving pressure, and statistically relates computed pressure measures to ball events and players' positions. Section 4 shows how computed measures are used in visually supported analysis, and then Section 5 explains how interesting subsets of game episodes and situations are interactively extracted. After a discussion of our work (Section 6), we conclude in Section 7. High resolution figures and a supplementary video are available at http://geoanalytics.net/and/dmkd2016/.

2 Related work

2.1 Analysis of movement data in football

To the best of our knowledge, data analysis was introduced to football in the early 1950s by Charles Reep[5] who proposed basic indicators such as pass to goal ratio. His methods have been used by professional clubs and published in scientific literature (e.g. [62]) and in mass media. In the 1970s, Zelentsov and Lobanovsky [81] started calculating so-called technical and tactical actions of all players of FC Dynamo Kiev and used their statistics as a basis for the team's training, tactics and strategy.

[4] http://www.bundesliga.de
[5] https://en.wikipedia.org/wiki/Charles_Reep

2.1.1 Analysis of game events

Nowadays game-tracking companies provide detailed statistics of all game events such as shots, passes, corners, ball recoveries, fouls, for each player and the team in general. Based on such statistics over multiple games, it is possible to understand a team's strategy, tactics and evaluate its performance. Several studies based on real data have been published in scientific literature.

Hirano and Tsumoto [37] analyze sequences of passes of varying length and discover some interesting pass patterns that may be associated with successful goals. Gyarmati et al. [34] collect statistics of motifs occurring in pass sequences of teams and cluster the teams by similarity of their motif sets to reveal patterns of passing strategies used by the teams. Clemente et al. [18] build a connectivity matrix containing counts of co-occurrences of pairs of players within the same offensive sequences of passes without losing possession of the ball.

Duch et al. [22] and Pena and Touchette [56] analyze pass graphs. Graph nodes represent players, with weights expressing their pass efficiency; edges correspond to passes between pairs of players weighted by their quantities. Their analysis of this network demonstrates that the more alternatives a team has for directing the ball to a shot, the better results that team achieves. Cintia et al. [17] use network centrality measures for analyzing a pass network, considering them from two perspectives: passes between players and passes between pitch zones. In a later paper [16], Cintia et al. calculate a single value - H indicator - that characterizes the passing behavior of a team. Their study shows that the H indicator is quite good at predicting the outcome of games.

Using event data from three seasons, Bojinov and Bornn [13] analyze statistics of pass disruptions by teams and relate them to other indicators such as controlling coefficient and shots made and conceded. They also study the spatial distributions of the disruption events for selected teams and for the league in general.

A SoccerStories system [57] automatically generates visual summaries of game episodes using a visual language with primitives representing game events such as long ball, turning the ball, square ball, long run, cross, corner and shot.

2.1.2 Analysis of trajectories

Further opportunities for analysis emerged due to the availability of detailed positional data covering movement of players and the ball. Thus, players' performance is often assessed using aggregate characteristics such as distance covered in different speeds at different stages of the game [20]. Various categories of player motions are identified [14] such as standing, walking, jogging, cruising (striding) and sprinting. Dividing player's trajectories into episodes of coherent motion extracts events that can be analyzed using the methods mentioned in the previous subsection. Rusu et al. proposed glyphs [64] for assessing multiple attributes of a player, comparing player's performance in multiple games, and comparing pairs of players. Specific glyphs have been proposed for goalkeepers [63].

Yue et al. [80] characterize behaviors of players and teams using time series of measures such as geometrical centers, radii, and expansion speeds. Kang et al. [42] proposed a method to extract kicks from trajectories and defined potentially reachable areas for players as cones in space-time. They also defined so-called catchable regions, where a player can potentially receive a pass, and competing and safe regions for players.

Several papers exploit the idea of generating Voronoi polygons [73] from positions of players (for example, Kim [45]). The area covered by a Voronoi cell can be interpreted as the control area of the player. Analysis of total areas over a game show that a team that occupies a larger proportion of the pitch has better chances to win. Taki et al. [69] extend this idea by computing dominant regions taking into account directions and speeds of players' movement. Horton et al. [38] use dominant regions for evaluating the quality of passes. Duarte et al. [21] analyze performance of football teams on the basis of general statistical indicators of team's convex hull such as area, circumference, length and width, which are calculated for the whole game and separately for periods of the match of pre-defined duration.

Websites and mobile apps of leading sport data providers often use heat maps of team's and player's presence in different parts of the pitch at different periods of the game, either in continuous (density map) or discretized (2D histogram) form. SoccerStories [57] build one-dimensional histograms of distributions of players along the horizontal or vertical dimensions of the pitch, allowing ball movements and passes to be related to the concentrations of players.

Lucey et al. [49] demonstrate that team strategy can be analyzed using only a discretized heatmap representation of the ball occupancy, irrespective of the positions and movement of the players. Comparison of the ball occupancy footprints of 20 teams in the English Premier League allows team behaviors to be discriminated, and alternative tactics to be identified for teams according to whether they are playing at their own stadium (home) or at the ground of an opponent (away). This approach permits assessment of whether a team's performance was within the expectation range in a selected match (i.e. the team played as usual) and enables outliers (atypical behaviors) to be identified.

2.1.3 Analysis of team formations

Knauf et al. developed a novel class of spatio-temporal convolution kernels [46] to capture similarities of collective movement of multiple players and the ball such as game initiations and scoring opportunities. The method identifies similar (in terms of spatio-temporal dynamics) configurations of sub-trajectories of players and the ball, thus helping to understand a team's strategy in given situations.

Wei et al. [76] built decision tree classifiers for identifying different stages of a game and roles of players. For the game stage detection, they segment a match in-play and stoppages. Further on, the stoppages are classified into out for corner, out for throw in, foul - free kick, and out for goal kick. The in-play segments are further divided into highlights and non-highlights. Highlights refer to all goal opportunities, both offensive and defensive. The classifiers have been built using a training set of manually labeled data for a single team, not considering their opponents. Laube et al. [47] discretized directions of player movement into 8 compass directions and then represented movement as a sequence of symbols. On the basis of this representation, they detect similar sequences performed by multiple players simultaneously and identify so-called trend-setters who change the direction of a team's movement.

Frencken et al. [25] study the variability of the distances between players and the team centroid. The study demonstrates that high values of this attribute are associated with collective defense actions and team reorganization during dead-ball situations. Moura et al. [53] calculate time series of the team convex hull area and the spread (based on pairwise distances between players). Fast Fourier transform of these time series is used to calculate median frequencies. The latter are interpreted as the speed at which players distribute and then compact their team formation during a match. Experiments show better efficiency during the first halves of games.

Several papers from Disney Research address a problem of reconstructing team formations and player roles from positional data. A role-based representation is dynamically updated for each time moment, capturing short-term context for both individual player and team analysis [12]. Based on positional data of a complete season of the English Premier League, 1,411 different formations were identified, each consisting of spatial distributions of positions corresponding to ten distinct roles of players. Role assignment is done for each player at each time moment, allowing identification of short- and longer-term roles and role swaps between pairs and larger groups of players. Further on, this approach enables identification of team styles that were used repeatedly in several games and detection of unusual behaviors [10]. Further analysis highlights differences between home and away team behaviors [11].

2.1.4 Analysis of derived features of team formations

Kim et al. [44] define a feature model for reducing the dimensionality and amount of trajectory data and producing data artifacts that are easier to analyze. Thus, they derive morphological properties such as width and depth of back-four (a defensive line of four players), and propose a morphological classification of features enumerating possible configurations of defenders, supporting analysis of their dynamics.

A similar approach is taken in the subsequent group of papers. Grunz et al. [28] analyze variants of game initiation by a team. By applying a self-organized map to $\{x,y\}$ positions of defenders, they identify a finite set of defense constellations. Next, equal-length sequences of constellations are grouped together for identifying repeatedly played game initiations. Following this idea, Perl and Memmert [59] reduce positions of players to tactical groups (e.g. offense or defense) that are further reduced to geometric formations (shapes) related to their centroid, irrespectively of their position on the pitch. Applying clustering to these geometric shapes identifies typical configurations used by a team. This work was further extended by Perl et al. [58] by calculating frequencies and success ratios for pairs of formations of two opponent teams.

2.1.5 Analysis of trajectory attributes

Janetzko et al. [40] support visual comparison of time series of characteristics of players, such as speed, acceleration, direction, cumulative distance, straightness, distance to next opposite, distance to ball, distance to own/opposite team center, and characteristics of teams, such as width and height of team's shape and quality of back-four formation (assuming that the team uses a tactical scheme with four defenders and their roles don't change during the game). An animated display enables viewing of selected subsets of the data. The paper also addresses an important problem of event annotation. Using partially annotated data, a classifier is built that is able to reconstruct additional candidate events. This work was further extended [68] for building a classifier that automatically extracts semantically-meaningful features such as short passing game, fast counterattack etc. Such a classifier helps to fix errors and fill gaps in manually annotated game event data. A further paper from the same group [65] proposes sophisticated visual search tools for identifying user-specified situations in tracking data. A visual query language allows either sketching a path line of a player or the ball or specifying areas of path start and end points. Further query conditions can be added through contextual menus. This query tool allows repeated combinations to be found in a game or across a series of games. Another kind of support for queries is provided in the system *Bagadus* for finding game events of interest such as "player X is in opponent's 18-yard box", "find all sprints faster than 10 mps" and "a striker is in a 3-meters distance from another player" [52]. Retrieved events are accompanied by corresponding videos.

2.1.6 Analysis of pass opportunities and scoring chances

Gudmundsson and Wolle [31] proposed a suite of techniques for analyzing potential pass opportunities of players in a given situation taking into account speeds and directions of players, for clustering of sub-trajectories of a player according to the similarity of geometric shapes (for example, detecting all starts of attack by a central midfielder), and for correlating movements with similar direction and shape of multiple players (e.g. simultaneous movement of several defenders).

Zhu et al. [83] combine the movements of players and the ball into a single aggregate trajectory to classify goal scoring events into categories. Lucey et al. [48] analyzed 10-second periods preceding 10,000 shots from a complete season in a professional league. They propose approaches to numeric estimation of scoring chances based on match context (open play, counter attack, corner, penalty, free kick, set piece), speed of play, shot location, defender proximity (including the number of defenders between the player who shoots and the goal and the distance to the nearest defender), and interaction with surrounding players.

2.2 Analysis of movement data in other sports

Football differs greatly from other kinds of team sports, being characterized by unique tactics, strategy and organization of play. Still, it is worth mentioning that there exist numerous studies on data analysis in other sports, for example, analysis of shots in ice-hockey [60], visual analysis of plays in American football represented by arc diagrams [55], representation of multiple aspects of ball-related events in rugby by glyphs [15], and comprehensive analysis of players performance and coordination in basketball [24]. A detailed review of these approaches specific to corresponding sport disciplines is beyond the scope of our paper. A survey [29] discusses common approaches in team sports analysis: playing area subdivision, identification of dominant regions of players and teams, network analysis, entropy, labeling game events and predicting future event types and locations, identifying formations, plays, and tactical group movements, and temporally segmenting a game. Performance measures are considered with a variety of offensive and defensive indicators. However, it is noted that appropriate performance indicators for football and methods for their derivation from position data are still missing.

Among the team sports, field hockey is deemed the most similar to football in terms of the rules, techniques, and tactics, including the use of pressure [51]. However, we are not aware of any research on pressure in field hockey. We expect that our approach to assessing and analyzing pressure in football may also be useful for field hockey.

2.3 Analysis of movement data in general

There exists a huge body of scientific literature on analysis of movement data in general, in application to land, sea, and air transportation, human mobility, animal ecology, eye tracking, just to name a few domains. The fields of movement analysis research include databases [33], data mining [27][82], geographic information science [30], and visual analytics [3]. Among the existing methods, there are generic methods suitable for a wide spectrum of application domains and more focused methods addressing particular kinds of movements (e.g., animal movement, city traffic, air traffic, movement of molecules, etc.) and application-specific analysis tasks.

Due to the large variety of moving objects, modes of movement, movement contexts, kinds of interactions between moving objects, etc., it is hardly possible to create a set of generic, universally applicable methods that can provide answers to all possible questions concerning movement. For football, both movements and questions are very specific, which necessitates development of special approaches. This explains the huge body of dedicated literature on football analysis. Still, many methods developed in other domains have potential to be adapted and applied to football data, providing new opportunities for analysis. Potentially relevant groups of methods are listed below; italic marks those that were used in our work reported in this paper.

- visualization and interactive exploration of spatial and temporal aspects of individual trajectories [43][74] and sets of trajectories [19][32][39],
- visual analysis of movement attributes along trajectories [67][71][79],
- *computational derivation of instant, interval and cumulative attributes of movement* [6],
- *detection of movement events*, such as stops, interactions between moving objects, significant changes, etc. [9][54],
- *extraction of events from trajectories* [5][6],
- classifying time intervals according to similarity of movement patterns based on presence at given places or moves between pairs of places [4][8],
- *deriving and visualizing spatial and temporal summaries of movement* [2][23][77][78],
- revealing relationships between movement and the environment (context) [5][50],
- *analysis of group movements* [7][61][72].

2.4 Interactive time filtering

Interactive exploration of spatio-temporal data often requires the data to be filtered based on the spatial locations, time references, values of attributes, and/or other conditions. Interactive filtering has been an important research topic in visualization and visual analytics since the pioneering work of Ben Shneiderman [66]. The book on visual analytics of movement [3] describes multiple filter types that are useful in exploration of spatio-temporal data, including selection of a time interval, several variants of spatial filters, attribute-based filter, trajectory segment filter, and cross-filtering between two datasets using references from one of them to items in the other one. Interactive cross-filtering is discussed in detail by Weaver [75].

There are two common approaches to interactive temporal filtering: by selecting a continuous time interval [1] or according to the positions of the time references within a time cycle [35][36]. These approaches are not sufficient for football analytics, where there is a need to select time moments and intervals based on values of time-dependent attributes, such as ball possession, status, and position on the pitch. It is important to be able to select all moments and intervals satisfying given query conditions and perform analysis on the data from these intervals. Query languages for temporal databases allow the selection of time-referenced data tuples the times in which belong to several disjoint time intervals satisfying some query conditions [26]. However, interactive tools enabling this kind of querying in the context of visual data exploration have been missing so far. In this paper, we fill this gap by introducing a novel interactive technique for temporal filtering. We call this type of filtering "time mask" as it hides (i.e., removes from consideration) the time intervals in which the query conditions do not hold. The remaining active (selected) time intervals may thus be separated by temporal gaps. The concept of time mask is similar to the concept of temporal element in temporal databases, which is defined as a finite union of n-dimensional time boxes [41].

The time mask filter can select data items from multiple disjoint and irregularly spaced time intervals of differing lengths. It is often unfeasible to consider each interval individually. A reasonable way to deal with the selected data items is to create and analyse various aggregates, which need to be dynamic: whenever the filter condition changes, the aggregates should be recreated by applying the aggregation operations to the newly selected subset of data items. The concept of "dynamic aggregators" has already been proposed [2]. A dynamic aggregator is a special object having references to a number of data records. It is able to check which records satisfy current filters and to derive certain statistical summaries from these records. The aggregators are responsible for the representation of the summaries on visual displays and for updating the view when the filters change.

2.5 Positioning of our work

In our paper, we address a football-specific research problem: how to measure, understand, and compare the use of pressing in defense behaviors of teams? As an approach, we introduce a numeric measure of pressure on opponent players and the ball. Besides calculating the pressure, we propose interactive techniques for detecting episodes and groups of episodes of interest (e.g. episodes of pressing), and visualizations for exploring player's and team's behavior in particular situations and, in aggregated form, in classes of game situations. From the perspective of visual analytics, our main contribution is the time mask filter, which is a novel flexible technique for interactive filtering and creation of statistical, spatial, and temporal summaries of multiple situations satisfying various query conditions.

3 Quantification of pressure

The pressing behavior of a defending team at each time moment can be seen as a combination of multiple instances of *pressure* relationships. In one instance, a defending team player (henceforth called *presser*) exerts pressure onto a *pressure target*, which is either the ball or one of the opponent players. Both the presser and the target can also be involved in other pressure instances, i.e., a target may receive pressure from several pressers, and a presser may exert pressure on several targets simultaneously.

There is no commonly adopted approach to quantifying pressure. So far, only one approach was proposed [70], in which the pressure on a player is calculated based on the distances of this player to the opponents and to the ball. This approach was judged too simplistic because it does not account for the directions the players are facing, which are intuitively perceived as being important [29].

We propose a new approach to quantifying pressure, in which directions are taken into account. It is hardly possible to reconstruct the directions faced by the players from trajectory data; however, face orientation may be less relevant than the directions in which the ball can be moved. Our approach is based on the following considerations. The aim of the attacking team is to move the ball towards the goal of the opponents. This can be done either by directly pushing the ball in the direction of the goal or by passing it to teammates, who can move it closer to the goal. The aim of a presser is the opposite: to prevent the ball from approaching the defended goal or to prevent an opponent player from sending or receiving a pass. Hence, the pressure exerted on a given target should be related to the directions from the target to the defended goal and to players of the attacking team (i.e., potential pass senders or receivers).

3.1 Pressure calculation

Let us call the direction from the pressure target towards the goal that is attacked or towards an attacking team player *threat direction*. Please note that there may be several threat directions for the same target. Let V^{threat} be the vector originating from the pressure target and oriented in one of the threat directions. The movement along this vector can most effectively be prevented or obstructed when the presser is positioned on V^{threat} in front of the target. The presser can also obstruct the movement along V^{threat} from a position aside of V^{threat}, but the distance to the pressure target in this case needs to be shorter. Even being behind the pressure target (with regard to the threat direction), the presser can still hinder the movement in the threat direction if the distance to the target is short enough. Hence, the maximal distance from which a presser can effectively pressurize a target depends on the angle $\Theta(V^{threat}, V^{TP})$ between the threat vector V^{threat} and the vector V^{TP} directed from the target towards the presser. When the angle Θ is zero, pressure can be exerted from a longer distance. As the angle increases (in the absolute value), the maximal distance for pressure decreases and reaches a minimum when $\Theta = \pm 180°$, i.e., when the presser is behind the pressure target.

This dependency can be geometrically represented by an oval where the pressure target is located at the narrow side whereas the wide side is oriented in the threat direction (Fig. 1). The exterior space around this shape can be considered as unsuitable for pressure on the target, i.e., we assume the pressure from any location beyond the oval to be zero. The interior of the shape can be called the "pressure zone". We specify the boundary of the pressure zone as a parametric curve with two parameters: front distance and back distance. The front distance D_{front} is the limit for exerting pressure when the presser is

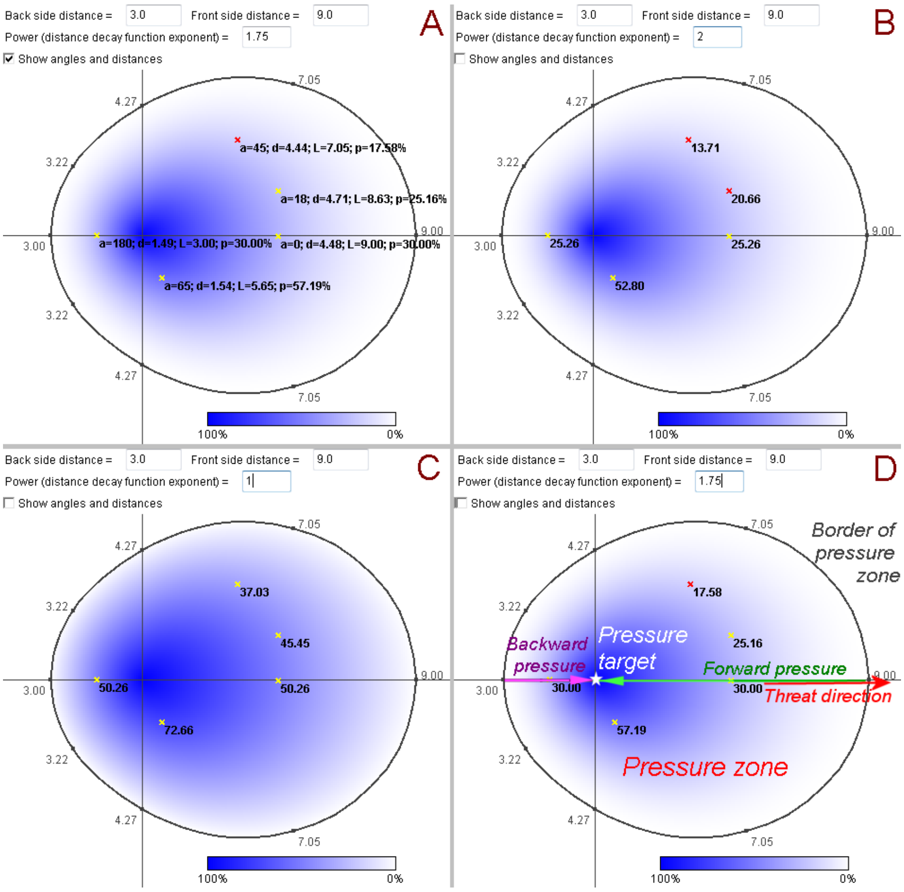

Fig. 1 The shape of a pressure zone and the distribution of the pressure levels. The variation of the pressure levels within the pressure zone is represented by varying the color saturation from fully saturated blue for 100% pressure to fully unsaturated (i.e., white) for zero pressure. The pressure levels in five selected points are shown by numbers. The images demonstrate the effect of using different exponents for the distance decay function: 1.75 (A and D), 2 (B), and 1 (C).

in front of the pressure target, i.e., when $\Theta(V^{threat}, V^{TP}) = 0$, and the back distance D_{back} is the limit in the case when the presser is behind the pressure target, i.e., when $\Theta(V^{threat}, V^{TP}) = \pm 180°$. The distance limits for the intermediate angles are determined by the following formula, which approximates an oval shape in polar coordinates (Θ, L):

$$L = D_{back} + (D_{front} - D_{back})(z^3 + 0.3z)/1.3, \tag{1}$$

where $z = (1 - \cos\Theta)/2$

The maximal theoretically possible pressure, taken as 100%, is when the presser is exactly in the location of the pressure target. For any point within the pressure zone, the pressure is estimated as

$$Pr = (1 - d/L)^q \times 100\% \quad (2)$$

where d is the distance of the point to the pressure target and L is the distance limit determined by formula (1). The exponent q regulates the speed of the distance decay, i.e., how fast the pressure decreases with distance.

The calculation of the pressure is illustrated in Fig. 1. Each image represents the pressure oval boundary of the pressure zone for the distance thresholds $D_{front} = 9$ m and $D_{back} = 3$ m, which were established in consultation with the football experts. The intersection of the horizontal and vertical axes in each image of Fig. 1 marks the position of the pressure target. The direction to the right is the threat direction (Fig. 1D). The numeric labels around the pressure oval show the values of the distance limit L corresponding to the angles $\pm 45°$, $\pm 90°$, and $\pm 180°$. The shading inside the curves represents the variation of the pressure levels within the pressure zone by proportional levels of color saturation, from fully saturated blue color for 100% pressure to fully unsaturated color, i.e., white, for 0% pressure. The pressure levels within the pressure zones in images A, B, and C have been calculated using different values of the distance decay exponent q: 1.75 in A, 2 in B, and 1 in C. It can be easily seen that, when q is higher, the decrease of the pressure with distance is faster.

To additionally facilitate comparisons, the pressure values in several sample points are shown by numbers. For the same points, the pressure values are lower when the exponent q is higher. Image A provides extended information for the sample points: besides the pressure level estimates shown as p=xx.xx%, the text also shows the size of the angle between the direction from the target to the selected point and the threat direction, denoted as a, the distance of the point to the target, denoted as d, and the distance limit for the angle a, denoted as L. Please note that the angles, distances, and distance limits are the same in all images but the pressure levels differ due to the differences in exponent q. Image D corresponds to the same value of q (1.75) as image A, but does not show the angles and distances at sample points.

The images in Fig. 1 are screenshots of a specially developed application that allows users to interactively change the values of the parameters D_{front}, D_{back}, and q and observe the impact on the estimated pressure levels. In particular, users can observe the distance decay effect of parameter q, i.e., how rapidly the pressure value decreases in different directions as the distance from the pressure target increases. The football experts were able to compare the different speeds of the distance decay for different values of q with their experience-based understanding of how the pressure level depends on the distance. Having experimented with the software, the experts came to the conclusion that an appropriate value for q should be between 1.5 and 2. Subsection 3.2 describes an experiment on verifying the expert-given estimations of D_{front}, D_{back}, and q by a computational experiment with real data.

As noted earlier, for one pressure target, there may be several threat directions, including the direction towards the goal and the directions towards the players of the attacking team, in particular, those that are located closer

Table 1 The games the data from which were used in the validation and analysis.

N	Date	Teams	Score
1	16/10/2015	1. FSV Mainz 05 - Borussia Dortmund	0:2
2	31/10/2015	SV Werder Bremen - Borussia Dortmund	1:3
3	20/11/2015	Hamburger SV - Borussia Dortmund	3:1
4	05/12/2015	VfL Wolfsburg - Borussia Dortmund	1:2

to the goal. The pressure $Pr(P,T)$ of a presser P onto a target T is computed for each possible threat direction, and then the maximum of these values is taken. The value of $Pr(P,T)$ is between 0 (no pressure) and 100% (maximal theoretically possible pressure).

Several pressers P_i can simultaneously exert pressure on the same target T. The total amount of pressure on T is the sum of the pressure amounts from all pressers: $Pr^{total}(T) = \sum Pr(P_i,T)$. The total pressure on a target can thus exceed 100%.

The calculation of the pressure level is applied to each time frame of the game when the ball was in play. Within a frame, the pressure on the ball is computed for each player of the defending team (i.e., the one without the ball). For each pair of players from the opposing teams, the pressure of the defending team player upon the player of the ball possessing team is computed. Then, for the ball and each player of the ball possessing team, we summarize the pressure amounts from all players of the defending team, and for each player of the defending team, we summarize the pressure amounts on all players of the opponent team. The computed pressure values can be visualized as shown in Subsection 3.3.

3.2 Verification of the parameters

We performed an experiment on checking the experts' judgments concerning D_{front}, D_{back}, and q against patterns that exist in real data. We used data from four away games of the team Borussia Dortmund (Table 1). The experiment was designed based on the following reasoning. One of the possible purposes for putting pressure on the ball is to re-gain ball possession. The pressure attempts that were made shortly before ball re-gain can thus be considered as successful. Let us find the players who could put the highest pressure on the ball shortly before ball re-gain, calculate their relative positions (i.e., the distances and angles) with respect to the ball and the threat directions, and look at the spatial distribution of these relative positions. Assuming that the likelihood of success increases with increasing the pressure, we can expect to observe the highest density of the "successful pressure" positions in a region where the players could exert the highest pressure. More generally, the overall density distribution of the "successful pressure" positions can be expected to correlate with the pressure levels attainable in different positions with respect to the ball.

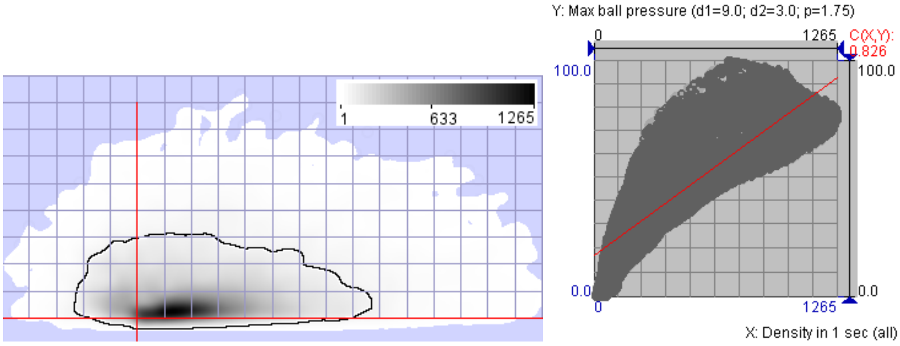

Fig. 2 Left: Density distribution of the relative positions of the most pressing defense players with respect to the ball in 1 second before ball re-gain in all 4 games. The isoline is drawn at the density level 63.3 (5% of the maximum). The inner region contains 85.5% of all positions. The intersection of the red axis lines marks the position of the ball. The grid with the cell size of 1x1 m shows the distances. Right: A scatter plot demonstrates a high correlation between the pressure levels and the densities of the pressing players' positions.

In the data from all four games, we found the time moments of ball possession change, excluding the situations when the ball was out of play before or after that. For the time intervals of length of 1 second preceding the possession change events, we found the defending team players who, according to their positions, could put the highest pressure on the ball. To identify these players, we applied the pressure estimation formulas (1) and (2) with relaxed parameter settings: $D_{front} = 15$ m, $D_{back} = 5$ m, and $q = 1$. The high values of D_{front} and D_{back} allowed us to include in the consideration also the players located beyond the pressure zone recommended by the football experts. The value $q = 1$ removes the effect of q in formula (2). Hence, the players were ordered according to the pressure potential of their positions regardless of the effect of q. For each time step in each of the 1-second intervals before the ball possession change, we recorded the player's position with the highest pressure potential, i.e., we recorded the distance d to the ball and the angle Θ with respect to the threat direction. In total, we extracted 31,827 such positions from all four games.

The density distribution of the extracted positions is shown in Fig. 2, left. The intersection of the horizontal and vertical axes (red lines) marks the position of the ball. The horizontal axis represents the threat direction, as in Fig. 1. Please note that the sign of the angle Θ with respect to the threat direction is not important for the pressure estimation because $\cos(\Theta) = \cos(-\Theta)$. Therefore, for building the density map in Fig. 2, we treated all angles as positive. To enable the estimation of distances, we created a grid with square cells of the size 1 x 1 m, which is put on top of the density map.

We see that the highest density of the "successful pressure" positions is reached within about 3 m in front of the ball, i.e., for the angles close to 0. By increasing the distance and/or the angle, the density monotonously decreases. This pattern is similar to the patterns of pressure visible in Fig. 1.

It should be taken into account that pressure on the ball is not the only possible reason for a change in ball possession. The latter may also be a consequence of an unsuccessful pass or goal attempt. Hence, not all of the 31,827 extracted positions may refer to cases of ball re-gain due to pressure on the ball. Finding a way to determine the true reason for each ball loss would require separate research, which is beyond the scope of this paper. Since we cannot select only relevant cases, we have to make approximate judgments using all extracted positions but keeping in mind that they include noise. It is safe to expect that the positions that are close to the ball and to the threat direction are more relevant, and that the relevance decreases (and, hence, the proportion of the noise increases) with increasing distance from the ball and the angle to the threat direction. From another perspective, it can be expected that relevant positions (i.e., positions of successful pressers) are more concentrated in zones from where the players could apply higher pressure upon the ball than in the areas that are less suitable for pressure, whereas the noise would be mostly distributed over the low density areas.

The map in Fig. 2 includes an isoline drawn at the density of about 5% of the maximal density value. The area enclosed in the isoline contains 85.5% of the extracted positions. Based on the previous reasoning, these positions are more relevant than the remaining positions located in the low density zone. The boundary of the higher density area reaches to about 9 m in front of the ball and about 2.5 m behind the ball. These distances match quite well the expert-given estimations of $D_{front} = 9$ m and $D_{back} = 3$ m.

To see the impact of the parameter q, we have applied formula (2) to the extracted positions using different values of q from 1 to 3 with a step of 0.25. As noted earlier, the pattern of the density distribution in Fig. 2 left is similar to the pressure variation patterns in Fig. 1. The scatter plot in Fig. 2 right shows a very high correlation between the density at a point (horizontal dimension) and the corresponding pressure level (vertical dimension). In this example, the pressure levels were obtained with $q = 1.75$, but plots built for the other values of q look very similar.

Table 2 contains the values of the Pearson's correlation coefficient between the position densities and the pressure levels computed with different values of q. The table columns correspond to the values of q and the rows to the games. The last row corresponds to the whole set of positions extracted from all games. The highest correlation value in each row is highlighted in bold.

In each row of Table 2 we see that the correlation for $q = 1.0$ is the lowest. It monotonously increases with increasing q to 2.0 or 2.25 and decreases as q increases further. The correlations for $q = 2.0$ and $q = 2.25$ are almost the same in the individual games and exactly the same for all games taken together. Based on the value of the correlation coefficient alone, a suitable value for parameter q would be 2.0 or 2.25. This does not perfectly match the experts' estimation that the value of q should be between 1.5 and 2.

However, there is another important criterion to be taken into account: how well the positions in the dense and sparse regions (i.e., zones of high and low relevance, according to the previous reasoning) are differentiated in

Table 2 Correlations between the density of the positions of the most pressing defense players in 1 sec before ball re-gain and the pressure on the ball calculated with different values of parameter q.

N	1.0	1.25	1.5	1.75	2.0	2.25	2.5	2.75	3.0
1	0.803	0.819	0.830	0.836	0.840	**0.842**	0.841	0.839	0.836
2	0.780	0.795	0.805	0.811	0.814	**0.815**	0.813	0.810	0.806
3	0.823	0.836	0.844	0.848	**0.850**	0.849	0.846	0.842	0.836
4	0.785	0.798	0.805	0.809	**0.811**	0.810	0.807	0.802	0.797
all	0.798	0.812	0.821	0.826	**0.828**	**0.828**	0.826	0.823	0.818

terms of the pressure levels estimated with different values of q. The lower values of q are better in respect to this criterion. It is obvious from formula (2) and also visible in the illustrations in Fig. 1 that the absolute values of pressure Pr decrease as q increases. For high values of q, the pressure estimates at positions located within the high relevance zone on medium distances from the ball become lower and thus closer to the pressure values on longer distances in the low relevance zone.

The value $q = 1.75$ is a good trade-off in respect to both criteria: while the correlation values are only slightly lower than for 2.0 and 2.25, the differentiation between the dense and sparse regions is better. Thus, the lower quartiles of the pressure values in the dense region for q =1.75, 2.0, and 2.25 are 29.2, 24.5, and 20.6, respectively, while the maximal values in the sparse region are 22.1, 17.9, and 14.4, respectively. For $q = 1.75$, 83.65% of all positions in the dense region have higher pressure levels than the maximum of the sparse region. For $q = 2.0$ and $q = 2.25$, these proportions are 83.56% and 83.59%, respectively. These differences, although not very high, support the experts' estimation of q. We would like to point out that the low differences for q =1.75, 2.0, and 2.25 with regard to both criteria indicate that results of pressure estimation are not very sensitive to the parameter settings. This robustness is a good feature of our approach.

Our experiment has helped the experts to see how their intuitive conceptions match real data. The choice of the parameter settings for pressure estimation was recently discussed at a one day workshop with participation of five football experts who brainstormed on various important concepts in football, including pressing tactics and the approaches to pressure quantification. It was thus finally agreed that the approach described in this paper is sound, and the values $D_{front} = 9$ m, $D_{back} = 3$ m, and $q = 1.75$ are appropriate. These settings have been used in all examples throughout the paper.

3.3 Examples of calculated pressure

On the basis of the data from the game between Borussia Dortmund (BVB) and VfL Wolfsburg (VfL) (see row 4 in Table 1), Figure 3 demonstrates a selected game episode with high levels of pressure on the ball and on the team

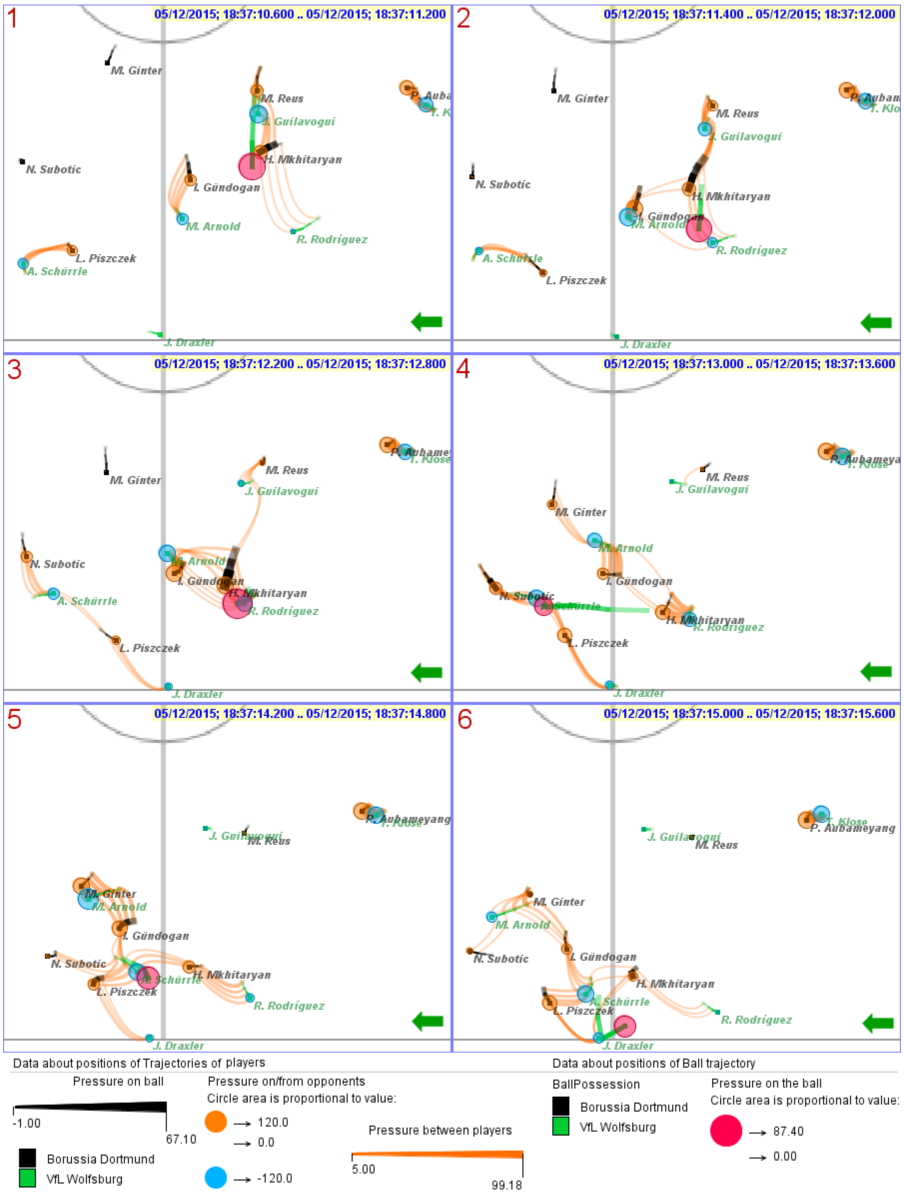

Fig. 3 Development of a game episode involving pressure on the ball and on players of the attacking team, namely, VfL Wolfsburg. The green arrows in the lower right corners of the images indicate the attack direction of VfL Wolfsburg. The circles in orange and blue represent positive and negative values of pressure between players, i.e., pressure exerted by players of the defending team (Borussia Dortmund) and pressure received by players of the attacking team. The sizes of the red circles represent the cumulative pressure on the ball. See the text for a full description of the figure.

possessing the ball (VfL). The goal of the defending team (BVB) is located on the left beyond the section of the pitch shown in the figure.

Explanation of the representation. Six chronologically ordered snapshots from an animated sequence represent situations in short time intervals (600 milliseconds length). The movements of the players during these intervals are represented by trace lines of two colors, green for VfL and black for BVB. The movement of the ball is shown by thicker trace lines painted in the color of the team possessing the ball, i.e., in green in this episode. The green arrows in the lower right corners of the images indicate the attack direction.

The circular symbols represent the pressure levels at the end of each time interval. The sizes of the red circles show the summary levels of pressure on the ball exerted by the players of the defending team (BVB). For particular players who exerted pressure on the ball, the levels of the pressure are represented by proportional widths of their trace lines. Circles in orange and blue are used to show positive and negative values of the pressure between players, where positive values represent exerted pressure and negative numbers represent received pressure. In this episode, the sizes of the blue circles show the summary levels of the pressure *received* by the players of VfL. The sizes of the orange circles show the summary levels of the pressure *exerted* by the defending players. Additionally, the pressure relationships between individual players from the opposite teams are represented by orange curved lines connecting the positions of the players at different time moments. The curvature of the lines increases in the direction from the pressure origin (i.e., a defending player) to the pressure target (i.e., an attacking team player). A sequence of such pressure lines between two players shows the dynamics of the pressure during the time interval represented in a map snapshots.

Interpretation and observations. VfL sends the ball to the left wing, where three players are available to receive it (snapshot 1). One of them (M.Arnold) is pressed by his opponent, therefore he moves forwards, giving space to R.Rodriguez (snapshot 2). BVB player H.Mkhitaryan moves towards R.Rodriguez, thus putting pressure on him (snapshot 3). Next, R.Rodrigez has to pass the ball to one of his teammates. Due to the configuration of the players on the pitch, he has only two feasible options, either to pass the ball in the direction of the BVB goal towards A.Schürrle (who is under pressure from N.Subotic) or to the left wing towards J.Draxler (pressed by L.Piszczek) (snapshot 4). A sequence of passes by VfL players moves the ball to the left wing and then to the back (snapshots 5-6). In this episode, the pressure generated by the BVB players forced VfL to slow down and eventually completely stop the attack by moving the ball towards the side of the pitch and then returning it to their defense.

It is worth looking not only at the players exerting and receiving high pressure but also at those who receive low or no pressure. Thus, in snapshots 4 and 5, J.Guilavogui (VfL) in the upper central part of the shown area was free from pressure and could have received a pass from A.Schürrle. Moreover, M.Arnold (VfL) moved forward, dragging I.Gündogan (BVB) with him and thereby allowing A.Schürrle to give a pass to J.Guilavogui. Had this hap-

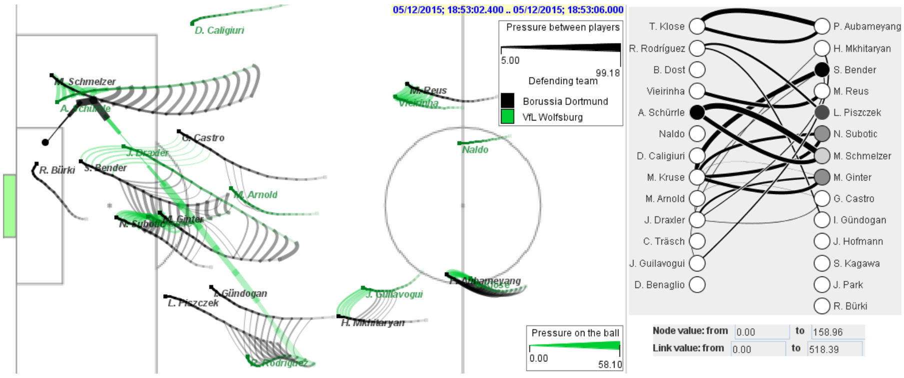

Fig. 4 Left: Player-to-player pressure is visually represented on a map by curved linear symbols. Right: For this game episode, a pressure graph shows the total amounts of pressure from the players on the ball (by node darkness) and on the opponents (by line widths).

pened, BVB's defense would have been compromised, since J.Guilavogui was free to penetrate it. Probably, because of the high pressure exerted by BVB, M.Schürrle did not see this option, which was fortunate for BVB. Hence, only the amounts of pressure and the effect do not tell everything about the quality of the defense. Although the coach of BVB might be happy that the ball was played back, he might also have serious concerns about the lack of pressure on J.Guilavogui.

An additional example. Figure 4 illustrates in more detail the pairwise relationships of player-to-player pressure. The snapshot shows movements and events that happened during a 3-second interval. A long ball pass was made in the direction of the VfL player A.Schürrle (upper left corner). The movement of the ball is represented by a trace line colored partly in green and partly in black; the line thickness is proportional to the pressure on the ball; the ball position at the end of the time interval is marked by a black circle. While A.Schürrle was running towards the ball, M.Schmelzer from BVB was running in parallel and pressurizing A.Schürrle. This pressure is represented by a series of pressure lines directed from the trace line of M.Schmelzer to the trace line of A.Schürrle. The pressure lines are painted in the color of the presser. The line thickness is proportional to the level of the pressure. It can be seen how the pressure level increases as the presser comes closer to the target. At a certain time moment, M.Schmelzer managed to overtake the ball and pass it to the goalkeeper of his team. The change of the ball possession is seen from the change of the coloring of the trace line of the ball from green (VfL's color) to black (BVB's color). Starting from the moment of the ball possession change, the directionality of the pressure forces also changed. Now players of VfL exert pressure on players of BVB. This is represented by green coloring of the pressure lines. Thus, after the change of the ball possession, A.Schürrle started to pressurize M.Schmelzer.

Please note also the dynamics of the pressure relationships for other pairs of players. In particular, the pressure between Vierinha (VfL) and M.Reus

(BVB), whose traces are in the upper right part of the image, emerging only when the ball possession turned over to BVB. The direction is from Vierinha to M.Reus. Before that, although M.Reus was close to Vierinha, there was no pressure because M.Reus was on the opposite side of Vierinha with respect to the positions of the attacked goal and the ball. After the ball possession change, Vierinha exerted pressure on M.Reus from the side of the ball, to prevent M.Reus from receiving a ball pass from his teammates. The football experts commented on this situation as follows: the pattern when a player of a defending team does not put pressure upon an opponent located nearby and immediately becomes covered by this opponent after a change of the ball possession often indicates that this player is not positioned well enough.

On the right of Fig. 4, the pressure generated during the demonstrated episode is summarized in a graph representation. The graph nodes correspond to the players of VfL (left) and BVB (right). The darkness of the node shading is proportional to the amounts of pressure exerted by the players on the ball. The edges of the graph represent amounts of player-to-player pressure. The edge curvature increases in the direction from the presser to the target, and the width is proportional to the amount of the exerted pressure.

Looking at visual displays of these and other game episodes in static and dynamic modes (using map animation), the football experts agreed that what is represented as pressure corresponds to their intuitive understanding of pressing and marking. The visualizations prompted their reasoning concerning the relationships between players' actions and the pressure they received or the need to put pressure on the ball or opponents. This provides certain evidence that our approach to the numeric expression of pressure is valid and can be used for analysis. Nevertheless, we also undertook some numeric analysis.

3.4 Validation

There is no ground truth for checking the validity of our approach to pressure quantification. Neither labeled data are available (i.e., data with “real” pressure values) nor can experts provide numeric estimates for the pressure levels for different pairs (P_i, T_j) in different game episodes. The approach can only be validated in indirect ways. Thus, we can examine how pressure values are related to the expected outcomes of the pressure. The aim of the pressure on the ball is to either re-gain ball possession or stop the ball approaching the goal. If the events of a change of possession (the defending team winning the ball) and the ball moving away from the goal are associated with high levels of the pressure on the ball (i.e., significantly higher than usual), this can be considered as indirect evidence of the validity of the calculated values. These events are also expected to be associated with higher than usual pressure on the players of the ball possessing team. Besides, the pressure on the players is expected to be higher when they are in the defending team's half of the pitch and when they are close to the ball. If these expectations hold for the calculated pressure values, this can be taken as an evidence of the validity of our approach to pressure quantification.

For the validation, we used the data introduced in Subsection 3.2 (Table 1). To test the validity of the computed pressure on the ball, we compared the mean pressure on the ball over a whole game with the mean pressure on the ball in two- and one-second intervals before the events of (a) ball possession re-gain, (b) ball going out of play, and (c) ball turning away from the defended goal. The events were extracted from the ball trajectory. The events of the types (a) and (b) are the points in which the ball possession or status changed. For the events of type (c), we computed for each ball position when the ball was in play the changes of the ball distance to the defended goal in the past one second interval and in the next one second interval. If the change in the past one second was negative (less than -1), i.e., the ball approached the goal, and the change in the next one second was positive (more than 1), i.e., the ball having turned away from the goal, the point was taken as an event of ball turn away from the goal.

Table 3 contains the mean values of the pressure *upon the ball* exerted by the players of Borussia Dortmund (D) and by their opponents (O) in the four games. The values are given for the whole game (w), for the 1-second intervals before the events of types (a) and (c), and for the 2-seconds intervals before the events of type (b). It can be seen that the pressure on the ball before all three types of events is notably higher than on average during a game. For the events (a) and (c), the highest mean pressure values occurred in 1-second intervals before the events, while for the events (b), the highest values were attained in 2-seconds intervals before the events (however, the values in 1-second intervals were only slightly lower). This can be explained by the fact that situations of the ball being out of play are announced by referees, who need some time to react to an event having happened. During this reaction time, the pressure decreases. The complete statistics, including the means, standard deviations, and position counts for the 2-seconds and 1-second intervals are provided in the supplementary materials at http://geoanalytics.net/and/dmkd2016/. According to the t-test based on these statistics, all differences of the mean pressure values attained before the events (a), (b), and (c) to the overall means for the whole games are statistically significant, with the P values being less than 0.0001. The association between the high computed values of the pressure on the ball and the events that are known to often happen due to high pressure gives evidence of the validity of our approach to pressure quantification.

Analogously to Table 3, Table 4 shows the mean values of the pressure exerted by the players of the defending teams *upon a single outfield player* of the team possessing the ball over the whole game and before the events (a), (b), and (c). The pressure on the players was higher before the events than on average, but the differences are not as high as for the pressure on the ball. This is easily explainable: before such events happen, much pressure is received by players who are close to the ball, but there are also distant players receiving little or no pressure. The mean pressure value per player is therefore lower than the pressure on the ball.

Table 3 Mean pressure upon the ball. N: game number, in chronological order; D: Borussia Dortmund; O: opponent team; (w): whole game; (a): 1-second intervals before the ball possession re-gain; (b): 2-seconds intervals before the ball going out of play; (c): 1-second intervals before the ball turning away from the goal.

N	D(w)	O(w)	D(a)	O(a)	D(b)	O(b)	D(c)	O(c)
1	29.7	34.7	59.0	55.9	58.0	50.5	38.8	40.9
2	32.9	20.6	52.1	49.4	44.8	44.4	39.4	26.6
3	33.0	27.9	47.2	58.4	43.7	41.2	36.0	35.3
4	28.9	22.3	54.4	47.5	41.8	29.1	36.0	28.4

Table 4 Mean pressure upon an outfield player of the ball possessing team exerted by the players of Borussia Dortmund (D) and their opponents (O) over the whole game (w) and in 1-second intervals before the ball possession re-gain (a), 2-seconds intervals before the ball going out of play (b), and 1-second intervals before the ball turning away from the goal (c).

N	D(w)	O(w)	D(a)	O(a)	D(b)	O(b)	D(c)	O(c)
1	21.8	23.7	29.5	27.7	31.5	28.1	24.3	24.4
2	23.1	19.2	28.7	26.1	24.4	23.7	26.6	19.8
3	25.1	23.3	29.9	32.9	27.9	29.1	28.1	26.1
4	21.0	21.4	27.3	27.7	28.3	24.6	22.5	25.2

Table 5 Mean pressure exerted by the players of Borussia Dortmund (D) and their opponents (O) upon a single outfield player of the ball possessing team. The columns show the means computed over the whole pitch, over the defending team's half, and over the circular zones around the ball with radii 15-25, 5-15, and 0-5 meters.

N	Whole pitch		Defended half		15-25 m to ball		5-15 m to ball		0-5 m to ball	
	D	O	D	O	D	O	D	O	D	O
1	21.8	23.7	29.5	27.0	19.2	19.9	30.1	31.1	40.9	45.6
2	23.1	19.2	32.0	24.7	19.8	19.6	29.7	23.5	47.8	26.8
3	25.1	23.3	31.7	29.4	21.3	23.0	32.0	30.2	51.7	38.2
4	21.0	21.4	25.8	30.0	20.7	19.1	29.3	27.7	37.9	33.7

As noted earlier, the pressure on a player is expected to be higher when the player is in the defending team's half of the pitch or when he is close to the ball. The computed pressure values shown in Table 5 correspond to these expectations. The mean values in the defending team's half and in the distances 5-15 and 0-5 m from the ball are significantly higher, according to t-tests, than the mean values across the whole pitch. The mean values in the distances 15-25 m from the ball are much lower than in the smaller distances from the ball.

Overall, we can be conclude that the statistical tests provide solid, though indirect, evidence of the validity of our approach to the quantification of pressure on the ball and on the attacking team players. It

4 Analysis of pressure behaviors of teams

Static and dynamic displays of game episodes with explicit representation of pressure forces and levels, as demonstrated in subsection 3.3, can support the analysis of activities of particular players and groups of players in specific situations. In the context of our research, these displays also played another important role: they allowed the football experts to check the plausibility of the results of the pressure computations against their knowledge and mental models of pressure and pressing behavior in football. In this way, the experts acquired trust in the validity of the approach. However, to gain understanding of the overall tactics of the teams, it is insufficient to look at particular episodes. It is necessary to summarize multiple game episodes into overall views presenting general features of teams' behaviors. To support the analysis of the overall tactics of teams' defense, computed pressure values are summarized statistically (as in subsection 3.4), spatially, and temporally. The summarizing is applied to subsets of time intervals selected according to ball possession and status and, possibly, other attributes.

To generate spatial summaries, we compute weighted densities of the positions of the players or the ball using the pressure values at these positions as weights. The computation is done as follows. A space covering raster (grid with small square cells) is created. Initially, all cells receive zero values. For each position selected for summarization, the corresponding pressure value is added to the current value of the cell containing this position. Furthermore, spatial smoothing is applied, so that a point contributes not only to the cell containing it but also to the surrounding cells within a user-chosen radius (bandwidth). The amount of contribution to the surrounding cells decreases as their distance from the point increases. The user may select between linear and quadratic functions (kernels) for determining the distance decay. In the resulting raster, the cells are filled with summarized pressure values. The spatial summary can be visualized by representing the values in the cells by variation of shading, as shown in the figures that follow. On top of the shading, selected levels of summarized values can be marked by isolines.

It should be noted that variation of the smoothing bandwidth and kernel function does not significantly affect the spatial patterns that can be perceived. A wider bandwidth leads to higher spatial abstraction and smoother patterns. With a smaller bandwidth, maps look more patchy but general patterns are still visible and are consistent with more abstracted patterns obtained with wider bandwidths. In our examples, we used bandwidth of 5 meters. The raster cell sizes were 0.5 by 0.5 meters. All figures were produced with the use of the linear kernel function. For the quadratic kernel, the images look very similar. Anyway, the specific values contained in cells of a raster are not important, only the spatial patterns of relative differences are meaningful.

4.1 Comparing tactical behavior of two teams in a single game

As an example, we take the game of Borussia Dortmund (BVB) against VfL Wolfsburg (VfL); see Table 1. In the course of the game, BVB was in defense (i.e., without the ball) during 46,612 time steps of length 40 millisecond (about 31 minutes in total) and VfL during 36,850 time steps (24.57 minutes). BVB was able to re-gain ball possession 170 times and VfL 161 times. The ball went out of play 64 times under VfL's possession (and, respectively, BVB's defense) and 57 times under BVB's possession. The ball turned away from the goal 251 times during BVB's defense and 191 times during VfL's defense. Statistical summaries of the pressure (Section 3.4) say that BVB put significantly higher pressure on the ball than VfL throughout the game in general and before the events of ball possession re-gain, ball going out of play, and ball turning away from goal. Concerning the pressure on the opponent players, BVB put higher pressure upon players positioned close to the ball and in situations before the ball went out of play, whereas VfL exerted higher pressure on the opponents in their own half of the pitch.

4.1.1 Temporal summaries of pressure

In Fig. 5, temporal summaries are provided in the form of two-dimensional time histograms where the horizontal dimension corresponds to time and the vertical dimension to ball possession; hence, a histogram has two rows corresponding to ball possession by two teams. The time axis is divided into bins (time intervals), in which statistical summaries (counts, sums, averages, etc.) for these time intervals are represented by proportional heights of bars. In Fig. 5, the bins correspond to the minutes of the game (the game consists of two halves with a break between them, which is clearly reflected in the display). The histograms A and B show the amounts of pressure on the ball summarized by the minutes, and the histograms C and D show the same for the pressure on the opponents. The upper row in each 2D histogram corresponds to ball possession by BVB and, hence, shows the pressure exerted by VfL. The lower rows correspond to ball possession by VfL and show the pressure exerted by BVB.

The 2D histogram in Fig. 5A shows that the amounts of pressure on the ball coming from BVB were usually higher than the pressure from VfL. In Fig. 5B, the same histogram appears in a mode in which the values are accumulated along the X-axis. In this mode, each bar represents the sum of all values from the preceding bars plus this bar, i.e., the cumulative statistics for the time since the beginning of the game till this minute inclusive. The differences between the amounts of pressure generated by the two teams are more prominent. Comparison may be further facilitated by showing the cumulative amounts of pressure from VfL on top of the bars showing the pressure from BVB. The pressure from VfL is shown in the lower row of the histogram by semi-transparent red bars, so that the heights of the bars of BVB and VfL can be easily compared. It can be seen that, starting from minute 20, the cumulative

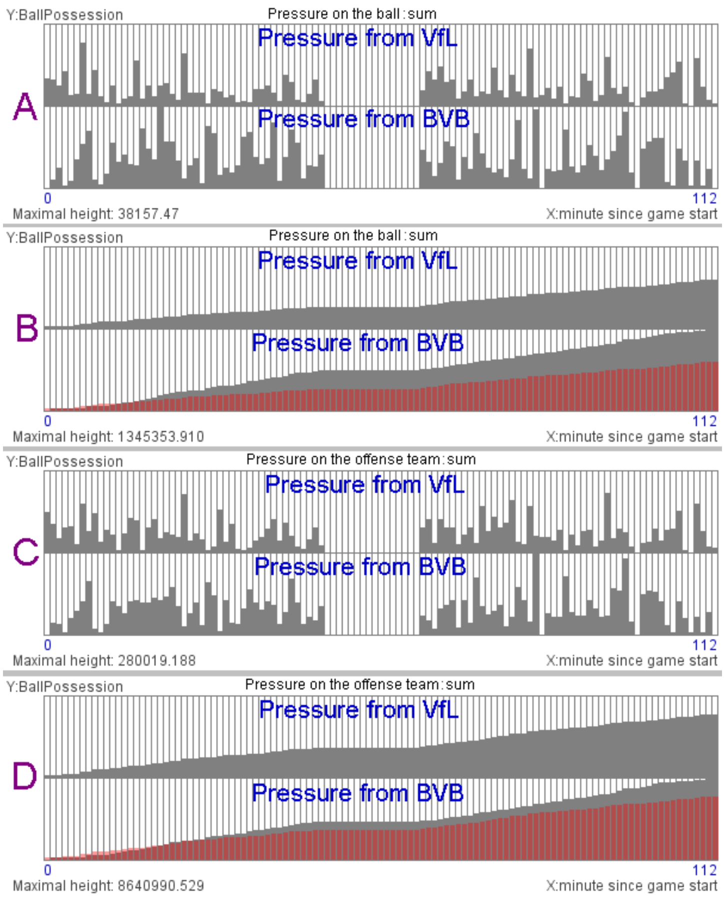

Fig. 5 Two-dimensional (2D) time histograms show the amounts of the pressure on the ball (A, B) and opponents (C, D) summarized by 1-minute time intervals. The horizontal dimension represents time. Each 2D histogram shows the pressure exerted by two teams, VfL in the upper row and BvB in the lower row. In histograms B and D, the amounts of pressure are accumulated along the time axis.

pressure generated by BVB grows faster than the pressure from VfL. By the end of the game, the difference is about 40%.

The histograms C and D are analogous to A and B but show the pressure on the opponents. We see that the differences in the pressure on the opponents are not as high as the pressure on the ball. The cumulative histogram (D) shows that by the end of the game the difference in the amounts of generated pressure was about 20%.

4.1.2 Spatial summaries of pressure

Coordinate transformation in the second half-game data. In the second half of the game, the teams exchange ends. To simplify joint analysis and comparison of data from both halves, we transform the data from the second half-game: we replace the x- and y-coordinates by the symmetric coordinates with respect to the vertical and horizontal axes of the pitch, respectively. As a result, the teams have the same sides of the pitch in both halves of the game. For the game of BVB against VfL, the goal of BVB is, after the coordinate transformation, always on the left and the goal of VfL is on the right.

Explanation of the representations. Figure 6 presents maps of players' position density and team formations, which are provided as a basis for comparison with the spatial summaries of pressure introduced later on. The density of the positions of the outfield players (the goalkeepers excluded) is represented by shading; darker shades mean higher density. The upper pair of maps corresponds to the time intervals of ball possession by BVB and the lower maps to ball possession by VfL. The left pair of maps represents the attacking teams and the right pair of maps shows the defending teams. On top of the density shading, the average positions of the outfield players in the first and second halves of the game are marked by lighter and darker dot symbols, respectively. The number of dots per team and game half may exceed 10 (the number of outfield players) due to substitutions. BVB made three and VfL two substitutions, all in the second half of the game. The arrows in the corners show the attack directions of the teams.

Figure 7 presents the *spatial summaries of the pressure* on the ball and players, which show the variation of the total amounts of pressure over the pitch. The upper and lower maps summarize the time intervals of the ball possession by BVB and VfL, respectively. The upper pair of maps in Fig. 7 shows the pressure exerted by VfL while BVB possessed the ball and the lower pair of maps shows the pressure exerted by BVB while VfL possessed the ball.

On the left of Fig. 7 are maps of the *pressure on the ball.* The shading shows the amounts of pressure in different areas of the pitch. Darker shades correspond to higher amounts of pressure. On top of this shading, gray lines show the free movements of the ball, i.e., when the computed pressure level was zero. On the right of Fig. 7 are maps of the *pressure on the players* of the ball possessing team. The absolute values of the summary pressure levels over space are not easily interpretable; it only makes sense to consider the relative differences over the territory within a map and between two maps. To support comparisons between the maps of the pressure on the players, several pressure levels are represented by isolines. Each isoline connects points with the same level of pressure. In both maps, we introduced isolines for the same levels of pressure with equal steps between them.

Interpretation and observations. Both maps on the left (pressure on the ball) show that the defending teams tended to put less pressure on the ball in the opponents' halves of the pitch. There are more gray lines, i.e., free ball moves, in the attackers' halves, which are marked with arrows, than in the opposite halves. The shading in the background shows that BVB put a lot of defending pressure on both sides of the pitch. As expected, we see low pressure close to the corners and high pressure in front of the goals. This corresponds to the typical corner situations, when the players try to pass long balls directly

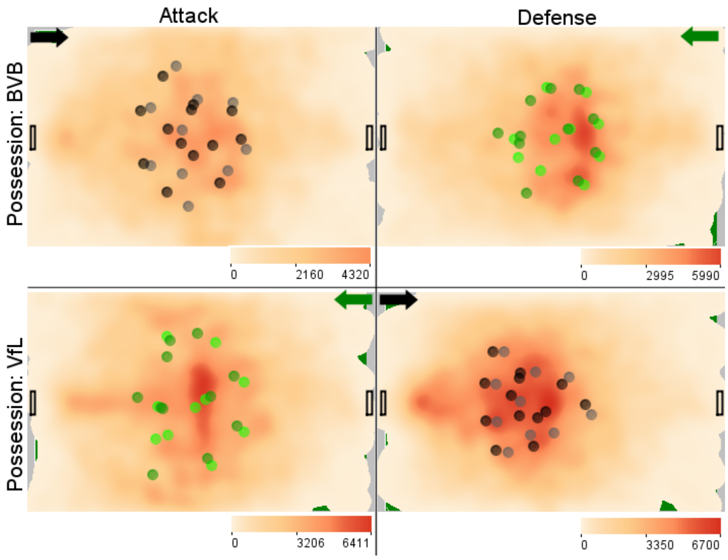

Fig. 6 Density maps represent the spatial distributions of the outfield players (goalkeepers excluded) of BVB and VfL during the ball possession by BVB (top) and VfL (bottom). Darker shading means higher density. The dot symbols show the average positions of the outfield players in the first (lighter dots) and second (darker dots) halves of the game. The arrows show the attack directions of the teams.

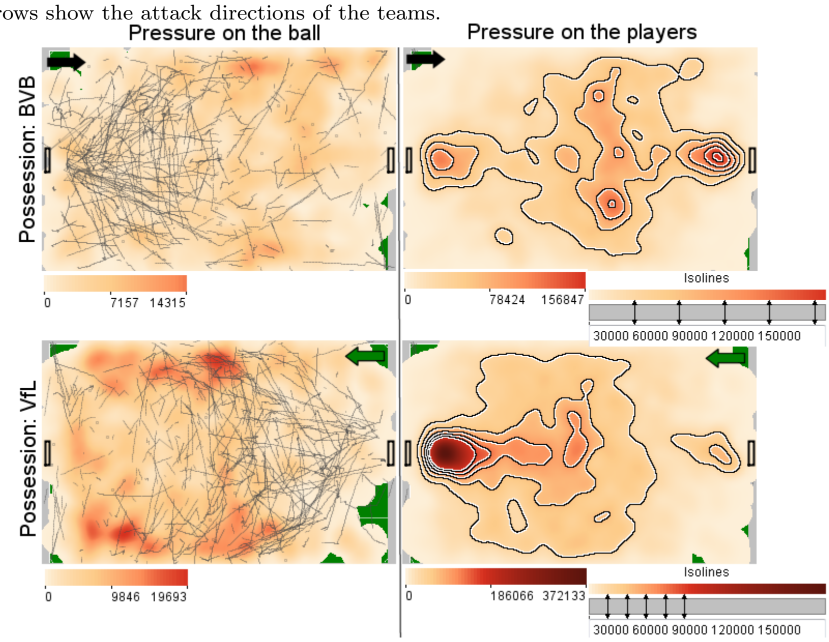

Fig. 7 Spatial summaries of the pressure on the ball (left) and on the players (right) during the ball possession by BVB (top) and VfL (bottom). On the left, the amount of pressure on the ball is shown by shading; darker shades mean more pressure. The gray lines overlaid upon the shading show the ball movements when there was no pressure. On the right, the amount of pressure on the players is shown by shading and, additionally, by isolines connecting points with the same pressure values.

from the corner to the area in front of the goal. Gray lines corresponding to such passes can be seen both in the upper and lower maps.

By comparing the upper and lower left maps, we see that BVB exerted much higher pressure on the ball during VfL's possession (lower map) than VfL did during BVB's possession (upper map). This observation is consistent with the statistical and temporal summaries discussed earlier. The pressure on the ball was especially high at the edges of the pitch. The areas of high pressure stretch over 75-80% of the pitch length, whereas VfL pressurized the ball mostly in their own half of the pitch.

Looking at the map of VfL's pressure on the BVB players (top right of Fig. 7), we see that the defense tactics of VfL was to create a kind of barrier of high pressure against the opponents across the pitch, to obstruct their both frontal and flank movements to VfL's goal. VfL also exerted much pressure on the BVB players in front of both goals. On the top left of Fig. 6, we see that the BVB players during their attack were distributed over a wide area, evidently, to find possibilities for approaching the opponents' goal along the wings. We found earlier that VfL did not exert too much pressure on the ball. The possible reason may be that BVB could rarely manage to overcome VfL's barrier of high pressure against the players, and thus the ball rarely appeared close to VfL's goal. We see in Fig. 7, upper left, that BVB's attempts to move the ball along the pitch edges were also strongly pressurized by VfL.

The lower left density map in Fig. 6 tells us that, when VfL possessed the ball, they tended to attack from the center of the pitch straight towards BVB's goal. The map on the bottom right of Fig. 7 shows us that BVB pressurized the players of VfL along this preferred line of attack, thereby forcing the attackers towards the wings. Through the high pressure on the VfL players along the X-axis of the pitch, BVB might also intend to obstruct passing of the ball from one side of the pitch to the other. The especially high levels of pressure of BVB against VfL in the area in front of BVB's goal can be explained by looking at Fig. 6, lower left: during their ball possession, VfL relatively often managed to get close to BVB's goal (it is seen from darker shading of the area in front of the goal). In such situations, BVB had to strongly pressurize their opponents in order to defend the goal.

The football experts were surprised to see a spot of very high pressure of VfL against BVB in the penalty area of BVB (bottom right of Fig. 7). Their guess was that BVB often gained the ball in front of their goal, e.g., after free kicks or intensive fighting for possession. At the time of gaining the ball, there were many VfL players in this area and, hence, high overall pressure. The visualization in Fig. 8 supports this interpretation. On the left, the dots show the locations of BVB's ball gain events, i.e., the positions of the ball when BVB re-gained possession. Indeed, many such events occurred in the penalty area of BVB. These events mostly happened after VfL's corners and crosses, which were, obviously, unsuccessful for VfL, as BVB managed to re-gain the ball. In this game, 9 corners were made by VfL and 5 by BVB, and the number of crosses inside the 18-yard line was 21 (VfL) against 8 (BVB); however, the accuracy of the crosses was 33% against 44%.

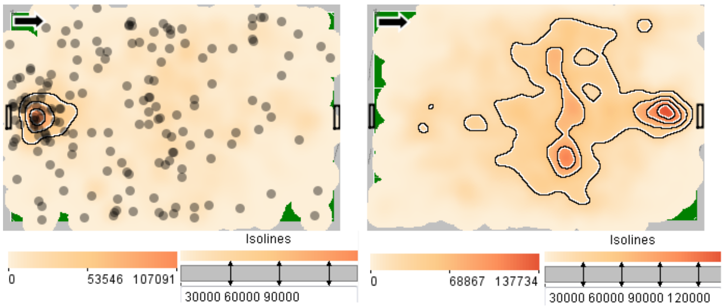

Fig. 8 Left: the pressure of VfL on BVB players in the first 3 seconds after BVB gained the ball. The dots show the locations of the events of ball re-gain by BVB. Right: the pressure of VfL on BVB players in the remaining time.

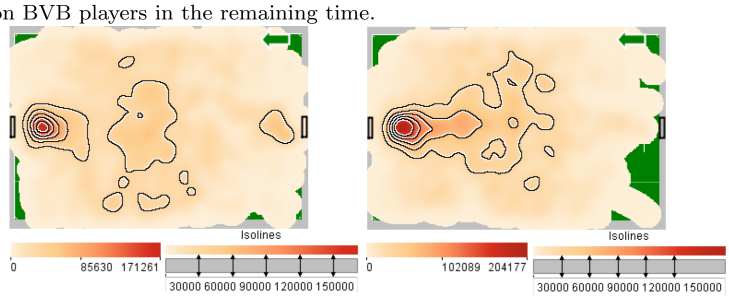

Fig. 9 The pressure of BVB on VFL players in the first and second halves of the game.

The background shading and isolines represent the pressure of the VfL on the BVB players in the 3-second intervals after BVB gained ball possession. The highest amount of pressure was in front of BVB's goal. The map on the right shows the distribution of VfL's pressure in the remaining time. We see that it was mostly in the middle of the pitch and at VfL's goal. To produce these maps, spatial summarization was applied separately to two collections of time intervals: first 3 seconds after each BVB's ball gain event and the remaining time of BVB's ball possession.

Figure 9 demonstrates that useful spatial summaries can be produced not only for the whole game but also by time intervals. As an example, the spatial distribution of the pressure of BVB upon VFL players in the first and second halves of the game are shown on the left and on the right of Fig. 9, respectively. We see that the prominent "ridge" of high pressure along the pitch length appeared only in the second half. In the first half, there was a "hill" on the left of the center of the pitch, which also evidences the tactics of BVB to deny the opponents the possibility to move in the center.

To summarize, while the tactics of VfL's defense was to create a barrier for preventing the opponents and the ball from approaching the goal, the tactics of BVB's defense was (1) to force the opponents to the wings by hampering their movements in the center of the pitch and (2) to exert high pressure on the ball for obstructing its progress along the pitch sides towards the goal and, whenever possible, re-gaining ball possession. Despite these tactics, VfL

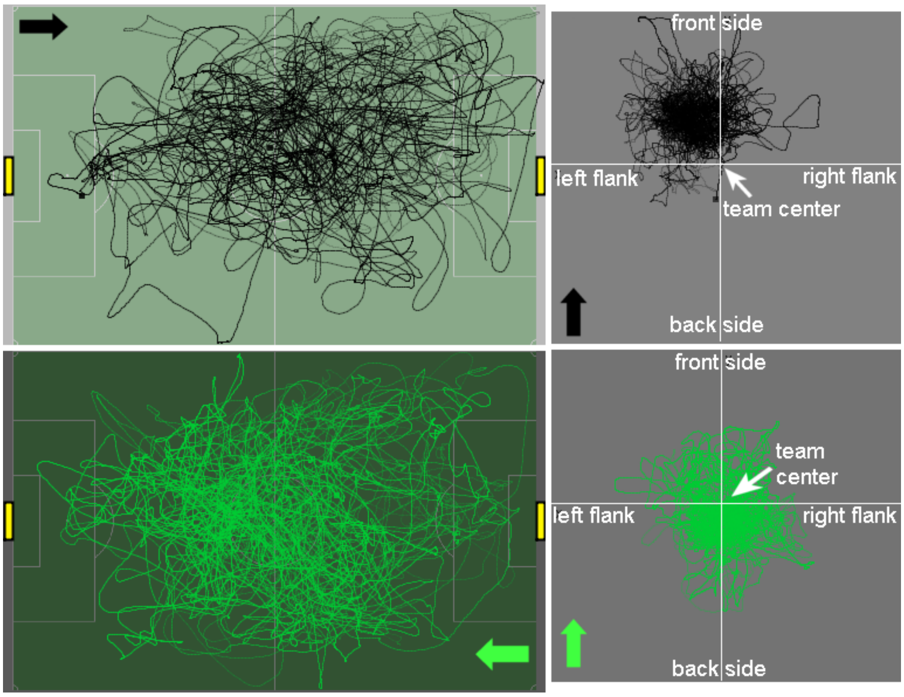

Fig. 10 Transformation of players' absolute positions to relative positions within their teams is illustrated by example of selected trajectories of two players, one from BVB (top) and another from VfL (bottom).

players frequently managed to reach the area in front of BVB's goal, where BVB had to exert much pressure on them for defending the goal and re-gaining the ball.

4.1.3 Distribution of pressure over team structure

Transformation of coordinates from absolute to relative. To consider how the pressure on the ball and on/from the opponents was distributed in relation to the players' positions within the teams, we apply a data transformation that maps the absolute positions of the players and the ball on the pitch into relative positions with respect to the centers of the teams and their directions of attack. The team center in each time moment is the average spatial position of all team members that are playing at this moment, except the goalkeeper. The transformation is illustrated in Fig. 10 by example of trajectories of two selected players from the opposite teams. On the left, there are two maps with the trajectory lines drawn according to the players' positions on the pitch. The two graphs on the right represent the same trajectories after the transformation.

In each graph, the crossing of the horizontal and vertical axes marks the center of the team (BVB in the upper graph and VfL in the lower graph).

The vertical axis is oriented from the goal of the respective team to the goal of the opponents, i.e., in the direction of this team's attack. This direction is represented by arrow symbols drawn in the lower left corners of the graphs. Hence, the upper part of a graph corresponds to the front side of the team and the lower part to the back side. The horizontal axis represents the lateral direction. The left and right side of a graph correspond to the left and right flank of the team. A player's position on the pitch in each time moment is transformed into the relative position with respect to the team's center and its medial and lateral axes. This can be seen as a transformation of the absolute, physical space into a relative, abstract "team space".

The graphs on the right of Fig. 10 show the relative spaces of BVB (top) and VfL (bottom). The trajectories of two selected players appear quite similar in the absolute space of the pitch (Fig. 10, left) but substantially differ when mapped onto the relative spaces of the respective teams (Fig. 10, right). The BVB player (black) was almost always in the front left part of his team, and the VfL player (green) mostly played in the center and center-back of his team.

The values of pressure that we derive from position data are associated with the positions of the players and the ball. After transforming the positions from absolute to relative, the pressure values can also be located in the abstract spaces of the teams. Spatial summarization of the pressure values can be done based on their positions in these abstract spaces, as shown in Figs. 11 and 12.

Explanation of representations. The shading and isolines show the distribution of the pressure on the ball (Fig. 11) and on the players (Fig. 12) in the team spaces. The visualization is analogous to the one applied in Fig. 7, right. The darkness of the shading is proportional to the total amount of pressure that occurred in different parts of a team. The isolines mark the same selected pressure levels across the four maps within each figure.

The pressure distribution is shown separately for the times of ball possession by each team. In each map, there is an arrow with the color indicating which team possesses the ball and the orientation showing the attack direction. When such an attack arrow is located in the lower left corner of a map and is oriented upwards, it means that this map shows the space of a team that possessed the ball, i.e., the attacking team. The shading shows how much pressure was generated by the members of the defending team (i.e., the one without the ball) in different positions with respect to the attacking team's center and attack direction (i.e., the direction towards the defended goal). When the attack arrow is drawn in the upper right corner of the map and is oriented downwards, it means that this map shows the space of a defending team. The shading shows how much pressure against the ball (Fig. 11) or against the opponents (Fig. 12) was generated by the defending team members in different relative positions within their team.

Interpretation and observations. In the upper two maps of Fig. 11, which show the pressure on the ball under BVB's possession, areas of high pressure can be observed in the front left quadrant of BVB's team space and on the right of the team center in VfL's team space. They tell us that the right midfield group of VfL put substantial pressure on the ball at the positions of

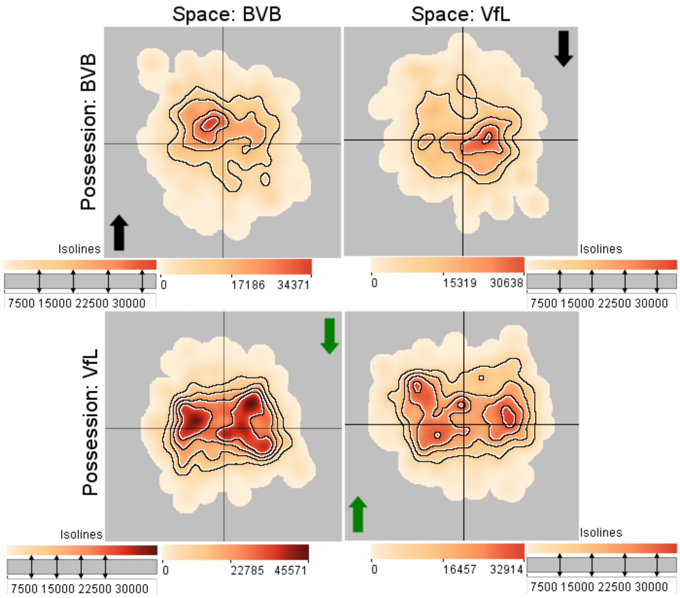

Fig. 11 Distribution of the amounts of pressure on the ball in the spaces of the teams BVB (left) and VfL (right) in situations of ball possession by BVB (top) and VfL (bottom). The arrows indicate the attack direction in each group of situations. The colors of the arrows indicate the attacking teams: black for BVB and green for VfL. The pressure on the ball is generated by the defending teams.

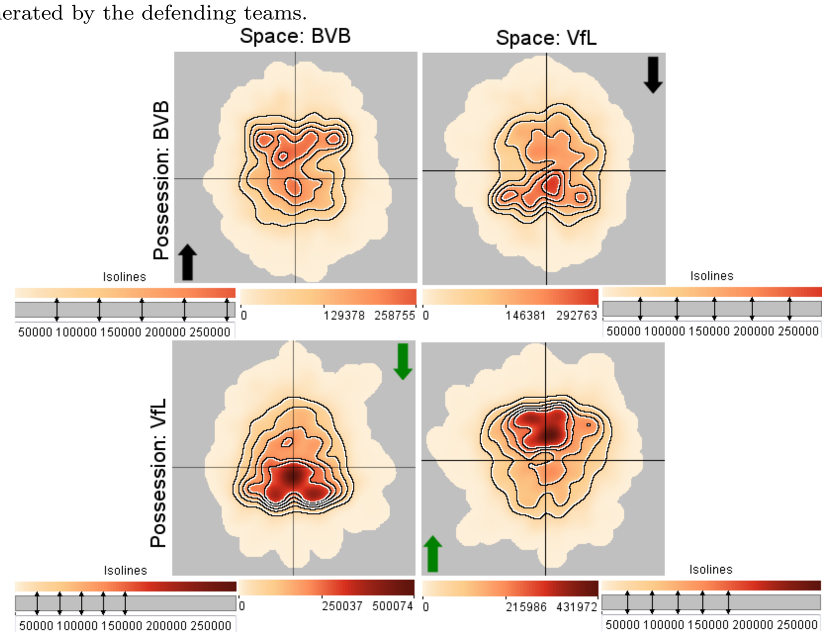

Fig. 12 Distribution of the amounts of pressure on the players of the ball possessing teams in the spaces of the teams BVB (left) and VfL (right) in situations of ball possession by BVB (top) and VfL (bottom). The arrows have the same meaning as in Fig. 11. The pressure on the attacking teams is generated by the defending teams.

the left midfield group of BVB. By looking at the average positions of the players in the team spaces, we can determine more specifically that the high amounts of pressure were, evidently, generated by J. Guilavogui and M.Arnold playing in the center and Vieirinha and D.Caligiuri playing on the right in VfL. The corresponding players of BVB were M.Reus and S.Kagawa.

Under VfL's possession (lower pair of maps in Fig. 11), BVB generated remarkably more pressure against the ball than VfL under BVB's possession. The highest pressure was concentrated in particular parts of BVB's team space. On the left, the highest amounts of pressure were generated by G.Castro in the first half and S.Kagawa in the second half. On the front right, the most active presser was H.Mkhitaryan. In the center, most pressure was generated by M.Ginter and I.Gündogan, and also by G.Castro in the second half. In the back right, the most active presser was L.Piszczek. The map of VfL's team space shows that the BVB pressurized the ball across the whole team space of VfL with emphasis on the VFL team sides.

In Fig. 12, which shows the distribution of the pressure on the players, we see that, under VfL's possession, BVB put strong pressure in the central triangle oriented towards the opponent's goal (Fig. 12, lower left). The pressure was concentrated at the average positions of M.Ginter (center), S.Bender (left back), and N.Subotic (right back). The target of the highest pressure was the front-central part of VfL (Fig. 12, lower right). This shows how BVB denied VfL the ability to play centrally.

Under BVB's possession, VfL's pressure was mostly generated by the back part of the team and targeted on BVB's attackers. This suggests that VfL's defense tactics was predominantly focused on halting BVB's attacks. VfL's cumulative pressure did not reach as high levels as the pressure from BVB's players. Similarly to the maps of VfL's pressure on the ball, we observe some asymmetry in the pressure distribution, with more focus on the left of BVB than on the right. The asymmetry of VfL's pressure on the ball and the opponents during BVB's possession may mean that BVB more often tried to attack from the left, or that VfL wanted to restrain particular BVB players from possessing the ball or hoped to quickly gain possession from these players. Our interactive tools allow us to identify that the most pressure was, evidently, directed against M.Reus, who played on the front left of BVB. The central part of VfL directed most pressure against H.Mkhitaryan.

By comparing the maps of BVB's pressure on the ball and the opponents under VfL's possession (lower left maps in Figs. 11 and 12), we observe that the players positioned in the front part and at the sides of BVB were predominantly focused on exerting pressure against the ball (Fig. 11) while the players in the back were more focused on pressurizing the opponents (Fig. 12). The central players contributed to both types of pressure. For VfL, we see a different division of the defensive responsibilities (cf. the upper right maps in Figs. 11 and 12). More pressure on the ball was exerted by the players positioned in the central right part of VfL, and more pressure on the opponents was generated by players in the left back of VfL. As with BVB, the central players were active in generating both kinds of pressure.

Hence, summaries of pressure distribution over team spaces provide additional insights into the teams' tactics by associating the pressure with team structure and the relative positions of team members.

4.2 Comparing tactical behaviors in two games

For comparison, we consider the game of Borussia Dortmund against Hamburger SV (HSV in short, symbolized by blue coloring in the following illustrations); see Table 1. The pair of figures 13 and 14 is analogous to Figures 6 and 7. Figure 13 shows the density distributions of the players' positions over the pitch and the average positions of the players during the ball possession by BVB (top) and HSV (bottom). The goal of HSV is, after the coordinate transformation, always on the right and the goal of BVB is on the left. The position densities during BVB's possession are much higher due to the longer total duration of ball possession by BVB (77,607 time steps = about 51.74 minutes vs. 54,928 time steps = about 36.62 minutes of HSV's possession).

Figure 14 shows spatial summaries of pressure on the ball and opponent players. The upper and lower images summarize the time intervals of ball possession by BVB and HSV, respectively. HSV used a defensive configuration in this game relying on counter attacks. The HSV team was compact in space and stayed close to their goal even during their possession of the ball (Fig. 13, bottom left). When not in possession of the ball, HSV applied very high levels of pressure on BVB's players across the whole width of the pitch in HSV's pitch half and especially in the vicinity of HSV's goal (Fig. 14, upper right). Such a behavior is often called "parking the bus". As a result, BVB's strikers had very little room to move. BVB were forced into making continual horizontal passes in the center of the pitch to vary the angle of attack (Fig. 14, upper left). However, on the flanks their ball possession was intensively pressed by HSV players. In contrast (Fig. 14, lower left), HSV often relied on long diagonal passes to the sides of the pitch center, expecting some chances for fast counter attacks. Several gray lines in the left side of the lower left map (meaning HSV's ball movements without the pressure from BVB) show the success of this strategy. This observation is supported by the unexpected result of this game, 3:1 in favor of HSV. Hence, the tactic of HSV was to sit deeply in front of their goal, prevent the opponents from approaching the goal, exhaust them by strong pressure, and apply counter attacks when managing to gain the ball.

Interestingly, some similarities can be observed between the spatial patterns of pressure in two games (cf. Figs. 7 and 14). The patterns of BVB's pressure against VfL (bottom of Fig. 7) look similar to the patterns of HSV's pressure against BVB (top of Fig. 14) after applying a symmetry transformation with respect to the vertical axis. The main difference is that the high pressure zone of HSV against BVB is strongly compressed on HSV's side of the pitch while the high pressure zone of BVB against VfL stretches over about 80% of the pitch length. At the same time, the pattern of BVB's pressure

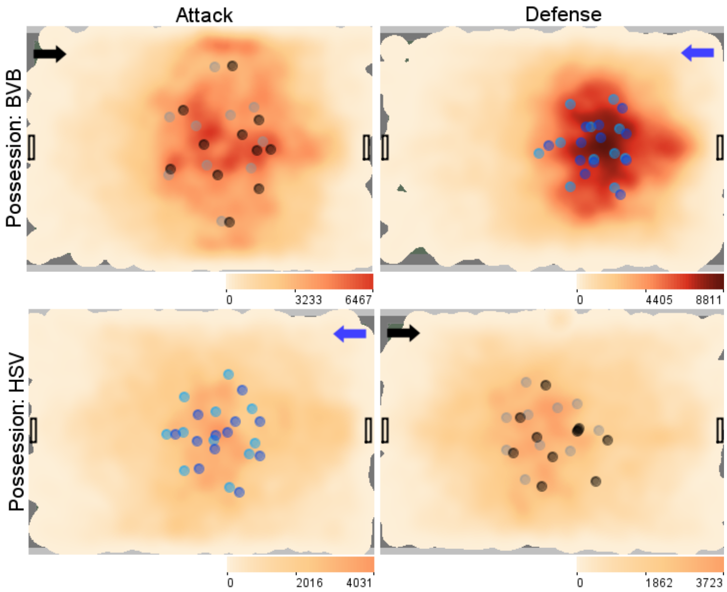

Fig. 13 Density of players' positions and team formations of BVB and HSV during the ball possession by BVB (top) and HSV (bottom). The lighter and darker dots show the average players' positions in the first and second halves of the game, respectively.

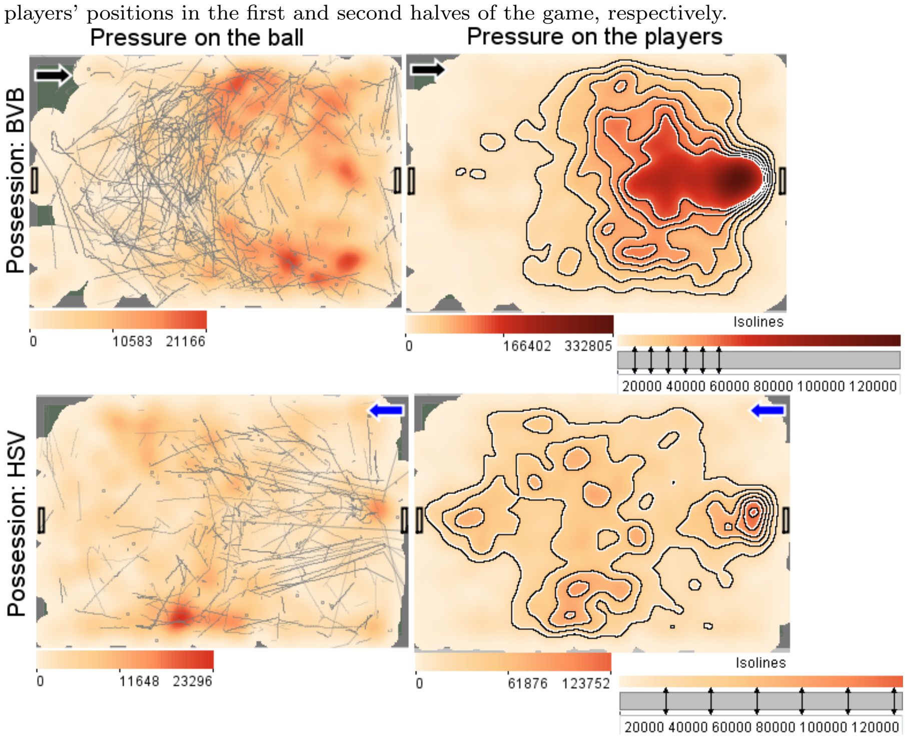

Fig. 14 Spatial summaries of the pressure on the ball (left) and on the players (right) during the ball possession by BVB (top) and HSV (bottom).

against HSV players (bottom right of Fig. 14) looks similar to the pattern of VfL's pressure against BVB players (top right of Fig. 7). We can observe similar barriers of high pressure. The difference is that VfL put their pressure barrier against BVB closer to the defended goal while BVB built their barrier against HSV in the center of the pitch. The cumulative pressure in the BVB barrier was lower, evidently, due to rare attacks of HSV. The similarity of pressure patterns across games and teams suggests that, possibly, there exists a set of "standard" defense tactics which are chosen by teams depending on the attacking style of their opponents.

Like Figures 11 and 12, Figures 15 and 16 show the distribution of the pressure on the ball and on the attacking players in the spaces of the teams BVB and HSV during each team's ball possession. There is a striking contrast between the upper and lower pairs of maps in both figures. BVB received much larger amounts of pressure during their possession than exerted during HSV's possession. The statistics in Section 3.4 tell us that the mean pressure levels on the ball and on a single attacking player in one time step were similar for both teams. Hence, the quantitative differences observed in Figs. 15 and 16 are due to the difference in the total durations of ball possession by the teams. BVB's possession was much longer, and so was HSV's defense.

HSV's pressure on the ball during BVB's possession stretches across the width of BVB's team space (Fig. 15, top left). Like VfL in the previously considered game, HSV put more emphasis on the front left quadrant of BVB's space. In this zone, there were average positions of M.Reus and S.Kagawa in the first half and of H.Mkhitaryan and G.Castro in the second half. In the distributions of the pressure against the players (Fig. 16), apart from the quantitative differences, there are some distinctions in the spatial patterns of the pressure. During HSV's possession, BVB distributed their pressure over the central area of HSV's space, with slightly higher concentration in the frontal part of it (bottom right). The distribution of BVB's pressure over their own team's space (bottom left) shows that the players in the team's front contributed to the pressure no less than the players in the back, i.e., the entire team was actively involved in pressurizing the opponents. This is different from the game against VfL, in which the pressure against the opponents was mostly applied by BVB's defensive players.

The pressure from HSV onto BVB (top left) was strongly focused on the BVB front line, i.e., against BVB's strikers. This pattern is somewhat similar to VfL's pressure against BVB (top left of Fig. 12), but the amount of HSV's pressure was much higher, and the pressure distribution was more symmetrical. In the HSV team's space (top right of Fig. 16), we see that the pressure was mostly generated by the players positioned in the back part of HSV, while the players in forward positions were less involved in pressurizing the opponents. A plausible interpretation is that the tactics of HSV was to keep their strikers possibly free somewhere behind the attack front of BVB, so that they could be ready to receive the ball in a case when HSV manages to gain it and then to be able to move with the ball towards BVB's goal.

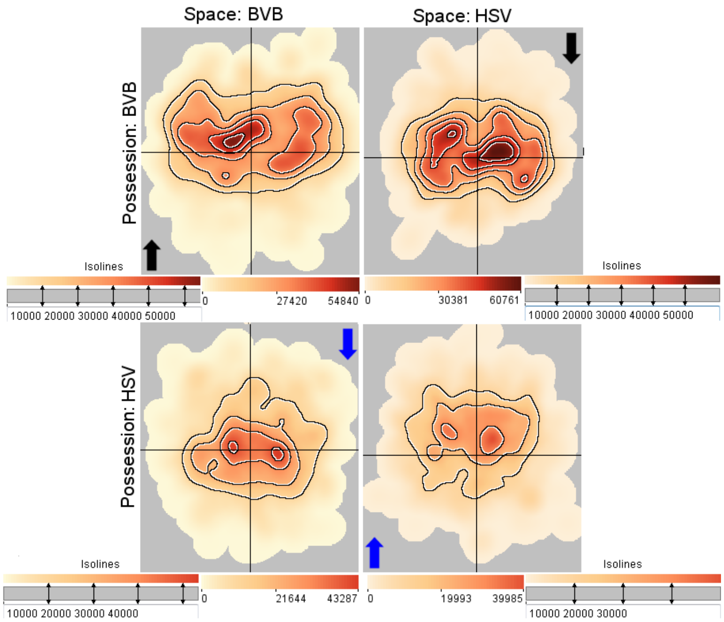

Fig. 15 Distribution of the amounts of pressure on the ball in the spaces of the teams BVB (left) and HSV (right) in situations of ball possession by BVB (top) and HSV (bottom). The arrows indicate the attack direction in each group of situations. The colors of the arrows indicate the attacking teams: black for BVB and blue for HSV.

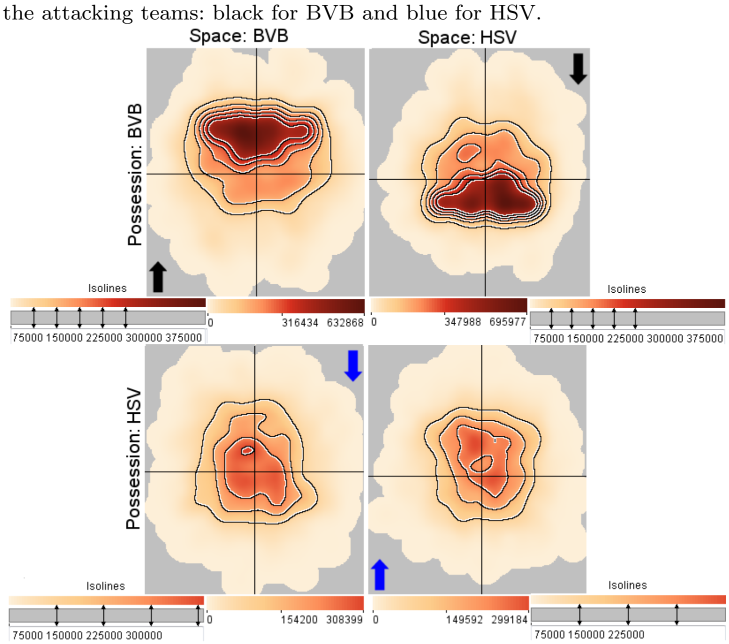

Fig. 16 Distribution of the amounts of pressure on the players of the ball possessing teams in the spaces of the teams BVB (left) and HSV (right) in situations of ball possession by BVB (top) and HSV (bottom). The arrows have the same meaning as in Fig. 15.

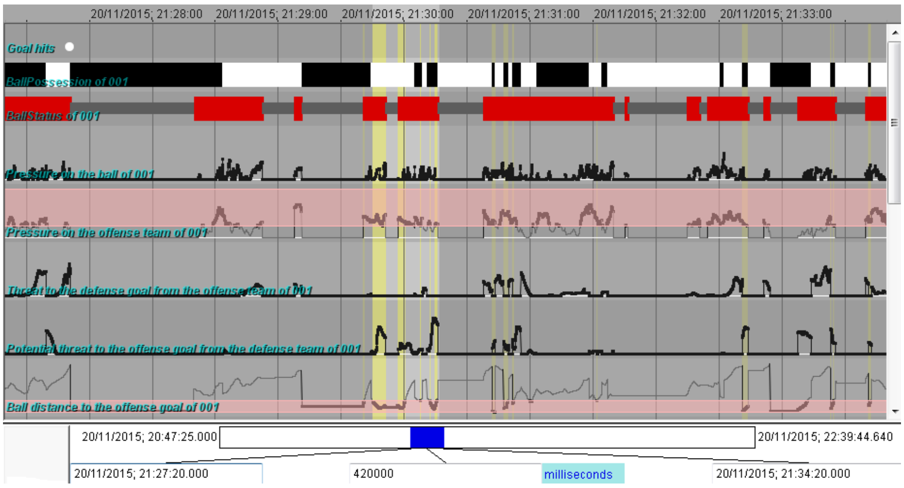

Fig. 17 An interactive display of time series of various attributes allows the user to set query conditions and select interesting game episodes for replaying.

Hence, by comparing the distributions of the pressure over the pitch and over the teams, we can reveal differences and similarities in defense tactics of the teams within and between games.

5 Interactive selection of game episodes and groups of situations

In analyzing data from a football game, the calculation of pressure indicators is applied to each time moment for which position data are available. The result is time series of the computed attribute values. Additionally, time series of values of other attributes may be computed, such as the ball distances to each team's goal, the widths and heights of the teams, the areas occupied by the teams on the pitch, and others. Our system provides the analyst an opportunity to select time series of interest for visualization in a time series display. An example is shown in Fig. 17. The horizontal axis of the display represents time. From an overview of the whole time span of the game, the user may zoom in to a shorter time interval to see the corresponding data in more detail. In Fig. 17, the display shows a selected interval of 7 minutes (420,000 milliseconds) length. The display is divided into horizontal bands given to different time series. A time series may consist of discrete events, such as shots on goal (the topmost band in Fig. 17), qualitative values, such as ball possession and ball status (the second and third bands), or numeric values (the remaining bands).

By mouse dragging within the display, the user can select a time interval for viewing statically or in animation. The selection works as a filter (called *time interval filter*): maps and other displays that are currently open represent only data from the selected interval. This refers also to the displays presenting statistical, spatial, and temporal summaries: the summaries are re-computed using only the selected data.

Within the time series display, the user can interactively set query conditions in terms of values of one or more attributes represented in the display. For qualitative attributes, the user can select particular values; for numeric attributes, value intervals can be selected. Within the display, the time intervals in which the attribute values satisfy the conditions of the query are marked by yellow background painting, which produces vertical stripes. This highlighting helps the user to find episodes of interest for viewing, such as episodes of high pressure on the ball and/or on players. Each episode can be selected individually by mouse dragging, which creates or redefines the time interval filter. In this way, we selected example episodes for presenting in Figs. 3 and 4.

Moreover, the selection of time intervals satisfying the query can itself be used as a filter. Unlike the previously mentioned time interval filter, which selects a single time interval, this type of filter, called *time mask filter*, may consist of multiple disjoint time intervals. Such a time mask can be created within the time series display and propagated to all other displays. In response, the displays show only data satisfying the time mask filter, i.e., belonging to the selected time intervals. Thus, on a map showing trajectories, only trajectory segments from the selected time intervals are shown. Displays presenting summaries re-compute the summaries using only data from the selected time intervals. In this way, in particular, we obtained the statistics presented in subsection 3.4 and the spatial summaries presented in Section 4.

It is possible to construct more sophisticated queries than just according to ball possession and status. Obviously, it is possible to select intervals based on the levels of pressure on the ball or the ball-possessing team, or pressure on/from particular players. It is also possible to set query conditions in terms of the time passed since the ball possession or status change and/or the time remaining to the next change of the ball possession or status. One example appeared in Fig. 8 showing the pressure on opponents in the first three seconds after losing the ball and in the remaining time of the opponent's possession. Another example is given in Fig. 18, which shows spatial summaries of the pressure on the ball by BVB in the game against VfL in the last 1-second interval before re-gaining the ball from the opponents. The images on the left and right correspond to the first and second halves, respectively, demonstrating that the time mask filter can be combined with the time interval filter. The images include the spatial summaries of the pressure on the ball and the movements of the ball in the selected intervals (i.e., the segments of the ball trajectory). In the upper pair of maps, the ball traces show where BVB regained the ball. We see that this often happened in the area in front of BVB's goal, but also in the other parts of the pitch. The shading shows the amounts of pressure generated in different parts of the pitch.

The lower pair of images shows the distribution of the pressure on the ball in the selected time intervals over the team space of BVB. On top of the pressure raster, the average positions of the players are shown. Please note that the average positions have been also computed only for the time intervals selected

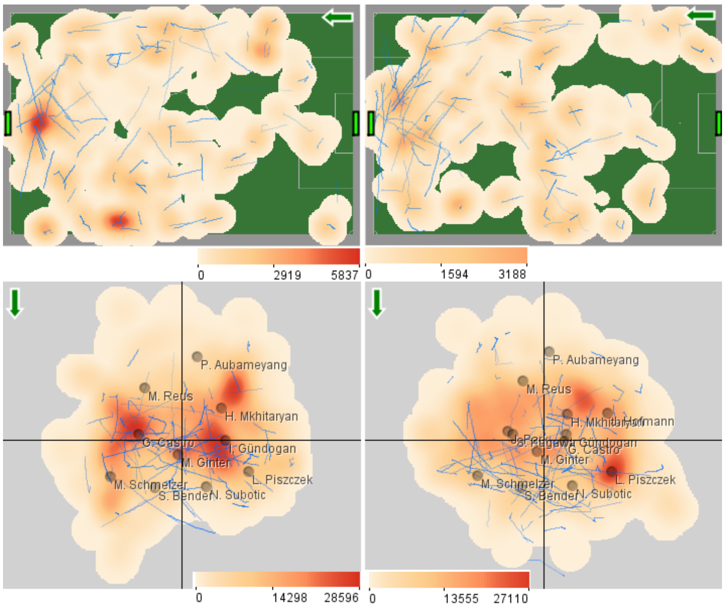

Fig. 18 Summaries of the pressure on the ball by BVB in the game against VfL in the last second before re-gaining the ball in the first (left) and second (right) halves of the game. The upper images show the distribution of the pressure in the pitch space and the lower images show the same in the team space of BVB, where the average positions of the BVB players in the selected time intervals are represented by labeled dots. The blue lines show the movements of the ball during the selected intervals.

by the combination of the time mask and time interval filters. We see that in the first half the highest amounts of pressure were generated near the average positions of G.Castro on the left, I.Gündogan on the right, and H.Mkhitaryan in the front right. In the second half, I.Gündogan had to be substituted in minute 54; therefore, he could not produce as much cumulative pressure as in the first half. The main source of pressure was L.Piszczek in the back right, while H.Mkhitaryan was active in the front right. Evidently, J.Hofmann, whose position is on the right of H.Mkhitaryan, also contributed to the pressure, but he appeared only in minute 82 as a substitute for P.Aubameyang.

Both types of time filtering can be combined with selection of particular players, for whom statistical, spatial, and temporal summaries of their movements, pressure they produced, and pressure they received can be obtained. Hence, the combination of the pressure computation, interactive filtering tools, data summarization tools, and interactive visual displays can support sophisticated analyses and answering questions concerning pressing and marking behaviors: When? Where? and Who?

6 Discussion

At present, there is no theoretical foundation of football tactics and adequate formalization of the essential concepts. From the perspective of football analytics, the main contribution of our work is that we have made a first step towards formalization of the concept of pressure and proposed a reasonable approach to pressure quantification. The measures we have introduced help to reveal pressing and marking behaviors of teams and analyze these behaviors in space and time within a single game and between games. The football experts note that it would be good to account additionally for the movement directions and speeds of the players and the ball in calculating pressure. But even the simple approach that we currently apply generates useful information for further analysis. The experts acknowledge that the new measures allow interpreting real data according to the universally understood (yet ill-defined) notions of football. The measure of pressure corresponds very well to the mental model of a coach, and pressing is a key concept in football that every coach wants to improve.

Our proposed approach to analyzing pressing and marking behaviors relies on visual and interactive analysis rather than automated computational methods. The approach involves computations for numeric estimation of the strength of pressure and for generation of various summaries that are visualized for human analysis. The reliance on the human may be seen as a weakness. However, experts believe that football tactics cannot be fully understood just from computed numbers without visualization and application of human capabilities for pattern grasping and reasoning. Football analysts are used to visual representations and able to extract knowledge from them. The visualizations of game episodes enhanced with explicit representation of pressure forces and pressure levels were quite easily interpreted and well appreciated by the football experts involved in this research.

In this regard, the football experts have suggested that the visualizations of game episodes can be enhanced even further by highlighting those players of the attacking team who receive no or too little pressure from defenders but might be dangerous in a case of receiving the ball. An example can be seen in the episode presented in Fig. 3. It would be good if J.Guilavogui in this episode could be somehow highlighted as potentially dangerous (see snapshots 4 and 5). This requires automatic detection of potentially dangerous players, which, in turn, require computation of pass opportunities, e.g., using the approach proposed by Gudmundsson and Wolle [31].

The spatial summaries of pressure were unusual to football experts, who found them difficult to understand. However, after gaining an understanding, they acknowledged the information richness and usefulness of these images. They have noted, however, that in order to ensure that our visualization designs can be understood and used by coaches (in particular, to analyze whether the actual game conforms to the planned tactics), they need to be involved in the research.

The experts have also noted that the pressure computed from real data and visualized by our methods may be hard to match to the ideas of coaches concerning the desirable tactical behaviors of their teams. The images of "real world" pressure derived from data differ from "textbook" pressure because in the real world the teams react to the behavior of the opponents. Therefore, our displays may show pressure zones that were not planned and emerged as a consequence of the other team's behavior. This can make it difficult for coaches to compare real pressure with what was planned.

In Section 3.4, we have correlated measures of pressure with ball events (possession change, going out of play, and turning away from goal), the players' positions in the pitch, and their distances to the ball. The experts suggest, as a continuation of the work, to correlate pressure measures also with other performance indicators, such as shots on goal and fouls.

It is important to stress that the suite of tools we developed to support the analysis of pressure includes not only computations and visual displays but also flexible and powerful tools for interactive querying (Section 5), which provide countless opportunities for focusing the analysis on particular phases and situations of the game, arbitrarily selected time intervals, and particular players. Thus, Figure 18 demonstrates how it is possible to visually relate pressure on the ball to ball recoveries, and Figure 8 shows how it is possible to separately analyze and compare the pressing behavior of a team immediately after losing the ball and later on. However, the flexibility of the interactive query tools may be overwhelming for end users. To improve the usability, it would be good, after communication with coaches, to define a set of queries that are likely to be frequently used and make them straightforwardly selectable on demand.

7 Conclusion

Modern tracking technologies enable movements of football players and the ball to be captured in fine detail and high quality movement data to be collected with relative ease, thus creating great opportunities for analysis. However, for gaining insights into the behaviors and tactics of football teams, it is not enough to have good data. It is necessary to understand how the concepts characterizing behaviors and tactics are represented in data and how to transform data into knowledge about behavior and tactics. This task is challenging given that these concepts are not well defined. We made an attempt to bridge the gap between raw positional data and the concepts of pressing and marking, which are very important in modern football tactics. Considering pressing and marking behaviors as a composition of multiple instances of pressure exerted by defending players on the ball and on the opponents, our group of computer scientists and football experts devised a method to detect and quantify these pressure relationships based on the relative positions of the players, the ball, and the defended goal. Then, we designed visualization methods to allow analysts to see the pressure relationships in detail, the overall patterns of their distribution over space and time, as well as the distribution with respect to the arrangements of the players in the teams. We also designed a novel interactive query tool "time mask" allowing flexible selection of potentially interesting game episodes for detailed analysis and subsets of game situations with specific properties for analyzing in summarized form. This kind of time-based querying and filtering has not been considered so far in the research fields of visualization and visual analytics. Certainly, there is room for further improvements in all components of our approach: computational, visual, and interactive. However, we have made an important first step towards, first, gaining an understanding of how pressing and marking behaviors are materialized in real data and, second, enabling analysis of these behaviors based on real data.

References

1. W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer Science & Business Media, 2011.
2. G. Andrienko and N. Andrienko. A general framework for using aggregation in visual exploration of movement data. *The Cartographic Journal*, 47(1):22–40, 2010.
3. G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual analytics of movement*. Springer Science & Business Media, 2013.
4. G. Andrienko, N. Andrienko, S. Bremm, T. Schreck, T. Von Landesberger, P. Bak, and D. Keim. Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. *Computer Graphics Forum*, 29(3):913–922, 2010.
5. G. Andrienko, N. Andrienko, and M. Heurich. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, 25(9):1347–1370, 2011.
6. G. Andrienko, N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel. Scalable analysis of movement data for extracting and exploring significant places. *Visualization and Computer Graphics, IEEE Transactions on*, 19(7):1078–1094, 2013.
7. N. Andrienko, G. Andrienko, L. Barrett, M. Dostie, and P. Henzi. Space transformation for understanding group movement. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2169–2178, 2013.
8. N. Andrienko, G. Andrienko, H. Stange, T. Liebig, and D. Hecker. Visual analytics for understanding spatial situations from episodic movement data. *KI-Künstliche Intelligenz*, 26(3):241–251, 2012.
9. P. Bak, M. Marder, S. Harary, A. Yaeli, and H. J. Ship. Scalable detection of spatiotemporal encounters in historical movement data. *Computer Graphics Forum*, 31(3pt1):915–924, 2012.
10. A. Bialkowski, P. Lucey, G. P. K. Carr, Y. Yue, S. Sridharan, and I. A. Matthews. Identifying team style in soccer using formations learned from spatiotemporal tracking data. In Z. Zhou, W. Wang, R. Kumar, H. Toivonen, J. Pei, J. Z. Huang, and X. Wu, editors, *2014 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2014, Shenzhen, China, December 14, 2014*, pages 9–14. IEEE, 2014.
11. A. Bialkowski, P. Lucey, P. Carr, Y. Yue, and I. Matthews. Win at home and draw away: automatic formation analysis highlighting the differences in home and away team behaviors. *Procedings of MIT Sloan Sports Analytics*, 2014.
12. A. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan, and I. A. Matthews. Large-scale analysis of soccer matches using spatiotemporal tracking data. In R. Kumar, H. Toivonen, J. Pei, J. Z. Huang, and X. Wu, editors, *2014 IEEE International Conference on Data Mining, ICDM 2014, Shenzhen, China, December 14-17, 2014*, pages 725–730. IEEE, 2014.
13. I. Bojinov and L. Bornn. The pressing game: Optimal defensive disruption in soccer, 2016.

14. C. Carling, J. Bloomfield, L. Nelsen, and T. Reilly. The role of motion analysis in elite soccer. *Sports Medicine*, 38(10):839–862, 2008.
15. D. H. Chung, P. A. Legg, M. L. Parry, R. Bown, I. W. Griffiths, R. S. Laramee, and M. Chen. Glyph sorting: Interactive visualization for multi-dimensional data. *Information Visualization*, 14(1):76–90, 2015.
16. P. Cintia, F. Giannotti, L. Pappalardo, D. Pedreschi, and M. Malvaldi. The harsh rule of the goals: data-driven performance indicators for football teams. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1–10. IEEE, 2015.
17. P. Cintia, S. Rinzivillo, and L. Pappalardo. A network-based approach to evaluate the performance of football teams. In *Machine Learning and Data Mining for Sports Analytics Workshop, Porto, Portugal*, 2015.
18. F. M. Clemente, M. S. Couceiro, F. M. L. Martins, and R. S. Mendes. Using network metrics in soccer: A macro-analysis. *Journal of human kinetics*, 45(1):123–134, 2015.
19. T. Crnovrsanin, C. Muelder, C. Correa, and K.-L. Ma. Proximity-based visualization of movement trace data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 11–18. IEEE, 2009.
20. V. Di Salvo, R. Baron, H. Tschan, F. Calderon Montero, N. Bachl, and F. Pigozzi. Performance characteristics according to playing position in elite soccer. *International journal of sports medicine*, 28(3):222–227, 2007.
21. R. Duarte, D. Araújo, H. Folgado, P. Esteves, P. Marques, and K. Davids. Capturing complex, non-linear team behaviours during competitive football performance. *Journal of Systems Science and Complexity*, 26(1):62–72, 2013.
22. J. Duch, J. S. Waitzman, and L. A. N. Amaral. Quantifying the performance of individual players in a team activity. *PLoS ONE*, 5(6):e10937, 06 2010.
23. J. A. Dykes and D. M. Mountain. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Computational Statistics & Data Analysis*, 43(4):581–603, 2003.
24. A. Franks, A. Miller, L. Bornn, K. Goldsberry, et al. Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics*, 9(1):94–121, 2015.
25. W. Frencken, H. d. Poel, C. Visscher, and K. Lemmink. Variability of inter-team distances associated with match events in elite-standard soccer. *Journal of sports sciences*, 30(12):1207–1213, 2012.
26. S. K. Gadia. A homogeneous relational model and query languages for temporal databases. *ACM Transactions on Database Systems (TODS)*, 13(4):418–448, 1988.
27. F. Giannotti and D. Pedreschi. *Mobility, data mining and privacy: Geographic knowledge discovery*. Springer Science & Business Media, 2008.
28. A. Grunz, D. Memmert, and J. Perl. Tactical pattern recognition in soccer games by means of special self-organizing maps. *Human Movement Science*, 31(2):334 – 343, 2012. Special issue on Network approaches in complex environments.
29. J. Gudmundsson and M. Horton. Spatio-temporal analysis of team sports - a survey. *CoRR*, abs/1602.06994, 2016.
30. J. Gudmundsson, P. Laube, and T. Wolle. Computational movement analysis. In *Springer handbook of geographic information*, pages 423–438. Springer, 2011.
31. J. Gudmundsson and T. Wolle. Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, 47:16–27, 2014.
32. H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan. Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *Visualization Symposium (PacificVis), 2011 IEEE Pacific*, pages 163–170. IEEE, 2011.
33. R. H. Güting and M. Schneider. *Moving objects databases*. Elsevier, 2005.
34. L. Gyarmati, H. Kwak, and P. Rodriguez. Searching for a unique style in soccer. *arXiv preprint arXiv:1409.0308*, 2014.
35. M. Harrower, A. L. Griffin, and A. MacEachren. Temporal focusing and temporal brushing: assessing their impact in geographic visualization. In *Proceedings of 19th International Cartographic Conference, Ottawa, Canada*, volume 1, pages 729–738, 1999.

36. M. Harrower, A. MacEachren, and A. L. Griffin. Developing a geographic visualization tool to support earth science learning. *Cartography and Geographic Information Science*, 27(4):279–293, 2000.
37. S. Hirano and S. Tsumoto. A clustering method for spatio-temporal data and its application to soccer game records. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 612–621. Springer, 2005.
38. M. Horton, J. Gudmundsson, S. Chawla, and J. Estephan. Automated classification of passing in football. In T. Cao, E.-P. Lim, Z.-H. Zhou, T.-B. Ho, D. Cheung, and H. Motoda, editors, *Advances in Knowledge Discovery and Data Mining: 19th Pacific-Asia Conference, PAKDD 2015, Ho Chi Minh City, Vietnam, May 19-22, 2015, Proceedings, Part II*, pages 319–330. Springer International Publishing, 2015.
39. C. Hurter, B. Tissoires, and S. Conversy. Fromdady: Spreading aircraft trajectories across views to support iterative queries. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1017–1024, 2009.
40. H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. A. Keim, and O. Deussen. Feature-driven visual analytics of soccer data. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, pages 13–22. IEEE, 2014.
41. C. S. Jensen, J. Clifford, S. K. Gadia, A. Segev, and R. T. Snodgrass. A glossary of temporal database concepts. *ACM Sigmod Record*, 21(3):35–43, 1992.
42. C.-H. Kang, J.-R. Hwang, and K.-J. Li. Trajectory analysis for soccer players. In *Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06)*, pages 377–381. IEEE, 2006.
43. T. Kapler and W. Wright. Geotime information visualization. *Information Visualization*, 4(2):136–146, 2005.
44. H.-C. Kim, O. Kwon, and K.-J. Li. Spatial and spatiotemporal analysis of soccer. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 385–388, New York, NY, USA, 2011. ACM.
45. S. Kim. Voronoi analysis of a soccer game. *Nonlinear Analysis: Modelling and Control*, 9(3):233–240, 2004.
46. K. Knauf, D. Memmert, and U. Brefeld. Spatio-temporal convolution kernels. *Machine Learning*, 102(2):247–273, 2015.
47. P. Laube, S. Imfeld, and R. Weibel. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, 19(6):639–668, 2005.
48. P. Lucey, A. Bialkowski, M. Monfort, P. Carr, and I. Matthews. Quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *MIT Sloan Sports Analytics Conference*. MIT Sloan, 2014.
49. P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews. Assessing team strategy using spatiotemporal data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 1366–1374, New York, NY, USA, 2013. ACM.
50. P. Lundblad, O. Eurenius, and T. Heldring. Interactive visualization of weather and ship data. In *Information Visualisation, 2009 13th International Conference*, pages 379–386. IEEE, 2009.
51. C. Mitchell-Taverner. *Field Hockey Techniques & Tactics*. Human Kinetics, 2005.
52. A. Mortensen, V. R. Gaddam, H. K. Stensland, C. Griwodz, D. Johansen, and P. Halvorsen. Automatic event extraction and video summaries from soccer games. In *Proceedings of the 5th ACM Multimedia Systems Conference*, pages 176–179. ACM, 2014.
53. F. A. Moura, L. E. B. Martins, R. O. Anido, P. R. C. Ruffino, R. M. L. Barros, and S. A. Cunha. A spectral analysis of team dynamics and tactics in brazilian football. *Journal of Sports Sciences*, 31(14):1568–1577, 2013. PMID: 23631771.
54. D. Orellana, M. Wachowicz, N. Andrienko, and G. Andrienko. Uncovering interaction patterns in mobile outdoor gaming. In *Advanced Geographic Information Systems & Web Services, 2009. GEOWS'09. International Conference on*, pages 177–182. IEEE, 2009.
55. S. G. Owens and T. Jankun-Kelly. Visualizations for exploration of american football season and play data. In *1st Workshop on Sports Data Visualization, IEEE VIS*, 2013.

56. J. L. Pena and H. Touchette. A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904*, 2012.
57. C. Perin, R. Vuillemot, and J. Fekete. Soccerstories: A kick-off for visual soccer analysis. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2506–2515, 2013.
58. J. Perl, A. Grunz, and D. Memmert. Tactics analysis in soccer–an advanced approach. *International Journal of Computer Science in Sport*, 12(1):33–44, 2013.
59. J. Perl and D. Memmert. Net-based game analysis by means of the software tool soccer. *International Journal of Computer Science in Sport*, 10(2):77–84, 2011.
60. H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko. Snapshot: Visualization to propel ice hockey analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2819–2828, 2012.
61. K. Reda, C. Tantipathananandh, T. Berger-Wolf, J. Leigh, and A. Johnson. Socioscape–a tool for interactive exploration of spatio-temporal group dynamics in social networks. In *Proc. IEEE Information Visualization Conference (INFOVIS)*, 2009.
62. C. Reep and B. Benjamin. Skill and chance in association football. *Journal of the Royal Statistical Society. Series A (General)*, 131(4):581–585, 1968.
63. A. Rusu, D. Stoica, and E. Burns. Analyzing soccer goalkeeper performance using a metaphor-based visualization. In *Information Visualisation (IV), 2011 15th International Conference on*, pages 194–199. IEEE, 2011.
64. A. Rusu, D. Stoica, E. Burns, B. Hample, K. McGarry, and R. Russell. Dynamic visualizations for soccer statistical analysis. In *Information Visualisation (IV), 2010 14th International Conference*, pages 207–212. IEEE, 2010.
65. L. Shao, D. Sacha, B. Neldner, M. Stein, and T. Schreck. Visual-interactive search for soccer trajectories to identify interesting game situations. *Electronic Imaging*, 2016(1):1–10, 2016.
66. B. Shneiderman. Dynamic queries for visual information seeking. *IEEE software*, 11(6):70–77, 1994.
67. D. Spretke, P. Bak, H. Janetzko, B. Kranstauber, F. Mansmann, and S. Davidson. Exploration through enrichment: a visual analytics approach for animal movement. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 421–424. ACM, 2011.
68. M. Stein, J. Häußler, D. Jäckle, H. Janetzko, T. Schreck, and D. A. Keim. Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *ISPRS International Journal of Geo-Information*, 4(4):2159, 2015.
69. T. Taki and J.-i. Hasegawa. Visualization of dominant region in team games and its application to teamwork analysis. In *Proceedings of the International Conference on Computer Graphics*, CGI '00, pages 227–235, Washington, DC, USA, 2000. IEEE Computer Society.
70. T. Taki, J.-i. Hasegawa, and T. Fukumura. Development of motion analysis system for quantitative evaluation of teamwork in soccer games. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 815–818. IEEE, 1996.
71. C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2565–2574, 2012.
72. T. von Landesberger, S. Bremm, T. Schreck, and D. W. Fellner. Feature-based automatic identification of interesting data segments in group movement data. *Information Visualization*, 13(3):190–212, 2014.
73. G. Voronoï. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. deuxième mémoire. recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik*, 134:198–287, 1908.
74. C. Ware, R. Arsenault, M. Plumlee, and D. Wiley. Visualizing the underwater behavior of humpback whales. *IEEE Computer Graphics and Applications*, 26(4):14–18, 2006.
75. C. Weaver. Cross-filtered views for multidimensional visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):192–204, 2010.
76. X. Wei, L. Sha, P. Lucey, S. Morgan, and S. Sridharan. Large-scale analysis of formations in soccer. In *Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on*, pages 1–8, Nov 2013.

77. N. Willems, H. Van De Wetering, and J. J. Van Wijk. Visualization of vessel movements. *Computer Graphics Forum*, 28(3):959–966, 2009.
78. J. Wood, J. Dykes, and A. Slingsby. Visualisation of origins, destinations and flows with od maps. *The Cartographic Journal*, 47(2):117–129, 2010.
79. M. Wörner and T. Ertl. Visual analysis of public transport vehicle movement. *5th International EuroVis Workshop on Visual Analytics (EuroVA'12)*, pages 79–83, 2012.
80. Z. Yue, H. Broich, F. Seifriz, and J. Mester. Mathematical analysis of a soccer game. part i: Individual and collective behaviors. *Studies in Applied Mathematics*, 121(3):223–243, 2008.
81. A. Zelentsov, V. Lobanovsky, V. Tkachuk, and A. Kondratjev. Tactics and strategy in football. *Zdorovja (in Russian)*, 1989.
82. Y. Zheng and X. Zhou. *Computing with spatial trajectories.* Springer Science & Business Media, 2011.
83. G. Zhu, Q. Huang, C. Xu, Y. Rui, S. Jiang, W. Gao, and H. Yao. Trajectory based event tactics analysis in broadcast sports video. In *Proceedings of the 15th international conference on Multimedia*, pages 58–67. ACM, 2007.